

The total external length of the evolving Kingman coalescent*

Iulia Dahmer,[†] Götz Kersting[†]

June 20, 2016

Abstract

The evolving Kingman coalescent is the tree-valued process which records the time evolution undergone by the genealogies of Moran populations. We consider the associated process of total external tree length of the evolving Kingman coalescent and its asymptotic behaviour when the number of leaves of the tree tends to infinity. We show that on the time-scale of the Moran model slowed down by a factor equal to the population size, the (centred and rescaled) external length process converges to a stationary Gaussian process with almost surely continuous paths and covariance function $c(s, t) = \left(\frac{2}{2+|s-t|}\right)^2$. A key role in the evolution of the external length is played by the internal lengths of finite orders in the coalescent at a fixed time which behave asymptotically in a multivariate Gaussian manner (see Dahmer and Kersting (2015)). A coupling of the Moran model with a critical branching process is used. We also derive a central limit result for normally distributed sums endowed with independent random coefficients.

MSC 2000 subject classifications. 60K35, 60F05, 60J10

Key words and phrases. evolving Kingman coalescent, external length process, Gaussian process, coupling, critical branching process

1 Introduction and main result

In mathematical population genetics the Kingman coalescent is a classical model for describing the genealogies of populations. If their population size is equal to n , the n -Kingman coalescent can be graphically represented as a binary tree which starts with n leaves and spends an exponential time X_k with parameter $\binom{k}{2}$ having k branches. The inter-coalescence times are independent. When labelling the leaves of the tree by $1, \dots, n$ one can define the n -Kingman coalescent as a partition-valued process $\Pi = (\Pi_k)_{1 \leq k \leq n}$ started in the partition $\pi_n = \{\{1\}, \dots, \{n\}\}$ of $\{1, \dots, n\}$ into singletons with the property that, when it is in a state π_k , it jumps after the exponential time X_k to a state obtained by merging two randomly chosen blocks from π_k . This happens in the coalescent time direction, from the present to the past.

As time runs forwards the population evolves and its genealogy changes, giving rise to a tree-valued process known as the *evolving Kingman coalescent* ([23],[24]). We consider in this paper populations started at remote past, so to say at time $-\infty$, and driven by the Moran model. This is a stationary, continuous-time evolution model in which each pair of individuals from the population is picked at rate 1, at which moment one of the individuals dies and the other one gives birth to an offspring, see e.g. [11].

The evolving coalescent discloses features of the Kingman coalescent which are less visible in the static model, when one analyses the tree only at fixed times. It is worth noting that some aspects of the tree arise from the recent past and others from the more distant past and that for certain functionals of the tree it may not be clear which one of these contributions dominates. As we shall see, the influence of the recent and the distant past is reflected among others in different time-scales, namely the evolutionary time-scale and the generations time-scale. We come back to this in more detail below.

*Work partially supported by the DFG Priority Programme SPP 1590 “Probabilistic Structures in Evolution”. I. D. was partially supported by the German Academic Exchange Service (DAAD).

[†]Institut für Mathematik, Goethe-Universität, 60054 Frankfurt am Main, Germany.
dahmer@math.uni-frankfurt.de, kersting@math.uni-frankfurt.de

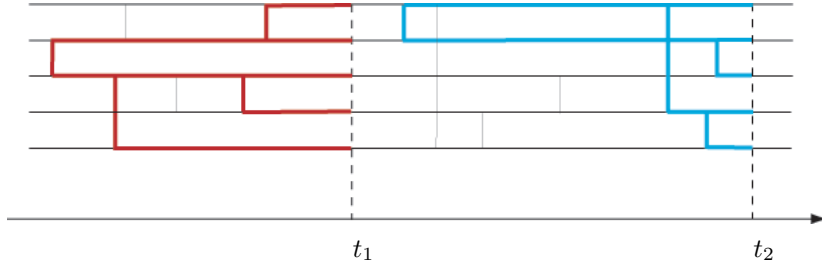


Figure 1: A realisation of the evolving Kingman coalescent at two different times for a Moran model with population size $n = 5$

Particular functionals of coalescent trees such as the total branch length (the sum of the lengths of all the branches of the tree) and the total external length (the sum of the lengths of the external branches) have been extensively investigated in the literature also due to their statistical relevance in population models including mutations. The mutation events are modelled as points of a Poisson process with constant rate on the branches of the coalescent tree. In the Infinitely Many Sites Model the total tree length controls the total number of mutations seen in the population whereas the total external length controls the number of mutations that affect only single individuals.

The Kingman coalescent is one of a big class of coalescent models which have been studied in the literature. The total length and the external length have been extensively studied for the Beta-coalescents. For the asymptotics in the static case we refer to [10], [17], [21], [3, 4], [2], [19], [9] and [6]. In the dynamic case, the evolving Beta($2 - \alpha, \alpha$)-coalescent was introduced and investigated for $\alpha = 1$ (the Bolthausen-Sznitman coalescent) in [25] and for $1 < \alpha < 2$ in [20]. In the latter case the processes of (centred and rescaled) total tree length (for $1 < \alpha < \frac{1}{2}(1 + \sqrt{5})$) and of total external length (for $1 < \alpha < 2$) converge in the sense of the finite dimensional distributions to stationary moving average processes with stable distributions. The time-scale is the generations time-scale for both the total length process and the external length process (which is obtained by slowing down time by a factor of $n^{\alpha-1}$). As we shall see, this is contrary to the Kingman case. Evolving coalescents have also been investigated from a different point of view, namely as evolving metric spaces in (e.g.) [8], [14], [15], [16].

In this paper we investigate the asymptotic behaviour of the total external length process of the evolving Kingman-coalescent. Observe that this process is *stationary*, a property that it inherits from the stationarity of the Moran model. The next theorem states our main result.

Theorem 1. *Let \mathcal{L}_t^n be the external length of the evolving Kingman n -coalescent at time $t \in \mathbb{R}$. Then, as $n \rightarrow \infty$*

$$\left(\sqrt{\frac{n}{4 \log n}} \left(\mathcal{L}_{t/n}^n - 2 \right) \right)_{t \in \mathbb{R}} \xrightarrow{f.d.d.} \mathcal{L},$$

where the limiting process $\mathcal{L} = (\mathcal{L}_h)_{h \in \mathbb{R}}$ is stationary, Gaussian, a.s. continuous, with mean 0 and covariance function

$$\text{COV}(\mathcal{L}_0, \mathcal{L}_h) = \left(\frac{2}{2+h} \right)^2, h \geq 0.$$

This convergence for the one-dimensional case can be read off from Janson and Kersting [18].

Let us contrast this result with the theorem of Pfaffelhuber, Wakolbinger and Weisshaupt [24], who investigated the process $(L_t^n)_{t \in \mathbb{R}}$ of total tree length of the evolving Kingman n -coalescent:

Theorem 2. *There exists a stationary process $L = (L_t)_{t \in \mathbb{R}}$ with paths in \mathbb{D} , the space of càdlàg functions equipped with the Skorokhod topology, such that*

$$L^n - 2 \log n \xrightarrow{d} L \quad \text{as } n \rightarrow \infty.$$

L_t has a Gumbel distribution for all t . Moreover

$$\frac{1}{t|\log t|} \mathbb{E}((L_t - L_0)^2) \rightarrow 4 \text{ as } t \rightarrow 0.$$

Theorem 2 deals with the original time-scale from the Moran model, that is the evolutionary time-scale. On the contrary, the time in Theorem 1 runs on the evolutionary time-scale *slowed down* by a factor of n . This is what we called the generations time-scale. Note that on the evolutionary time-scale reproduction events happen in the population with a rate of order n^2 . Therefore, for large populations, after a time of order $\frac{1}{n}$ each individual will have taken part in a number of reproduction events of order 1. It is not surprising that the external length has to be considered on this time-scale. It might be less obvious that the evolutionary time-scale is the appropriate choice for the total length.

We use the notation t for the time on the evolutionary time-scale and

$$h = h(n) := \frac{t}{n},$$

for the time points of the generations time-scale.

Note also that the limiting process \mathcal{L} is almost surely continuous, whereas L is almost surely made up of jumps. This reflects the fact that the total length experiences big jumps (at the times when old families become extinct). For the external length such extremal events do not come into play. This is also reflected in the type of limiting distributions. Let us point out that the process L is not a semimartingale ([7]), which so far is an open problem for the process \mathcal{L} .

In the rest of this introduction we outline the proof of our theorem. Due to stationarity it suffices to analyse the dynamics of the external length on a time interval $(0, h)$ with $h > 0$.

For describing the dynamics of the total external length of the Kingman coalescent we recall the notion of branches of order i introduced in [5]. In the coalescent tree each branch is situated above a subtree. If this subtree has i leaves, one says that the branch is *of order i* . If $i = 1$ the branches are external and if $i \geq 2$ they are internal. We denote by $\mathcal{L}_t^{n,i}$ the total length of order i (the sum of the lengths of branches of order i) in the n -coalescent belonging to the population alive at time t .

Let us now look at the dynamics of the total external length. On the generation time-scale each pair of individuals takes part in a reproduction event at rate $\frac{1}{n}$. As long as no reproduction event takes place, the external branches of the coalescent tree grow linearly and hence the external length grows at rate 1. At the time ζ of a reproducing event, the total external length process has a jump. The length of the branch b_1 corresponding to the individual that dies and the length of the branch b_2 corresponding to the individual that reproduces are subtracted from the total external length. The branch b_1 is removed from the tree, whereas b_2 becomes an internal branch of order 2. Typically, by the removal of b_1 , a branch that was internal (of order 2) at time $\zeta -$ becomes part of an external branch at time ζ (the exceptional case is when b_1 and b_2 stem from the same branching event). The length of this branch is then added to the total external length. Note that this piece may become internal again later on due to new reproducing events. If such an internal branch is part of an external branch at a future time h we say that the branch is *free* at time h . An example is given in Figure 2.

From the tree at time 0 not only internal branches of order 2 may be free at time h , but also internal branches of orders greater than 2. The lengths of the free branches become part of the external length at time h and hence $\mathcal{L}_h^n = \mathcal{L}_h^{n,1}$ can be written as

$$\mathcal{L}_h^n = \sum_{i=1}^{n-1} \Lambda^{n,i} \mathcal{L}_0^{n,i} + I_{0,h}^n, \quad (1)$$

where by $\Lambda^{n,i} = \Lambda_{0,h}^{n,i}$ we denote the (random) proportion of the length $\mathcal{L}_0^{n,i}$ of order i that is free by time h and by $I_{0,h}^n$ the external length of the coalescent at time h arising in the time interval $(0, h)$.

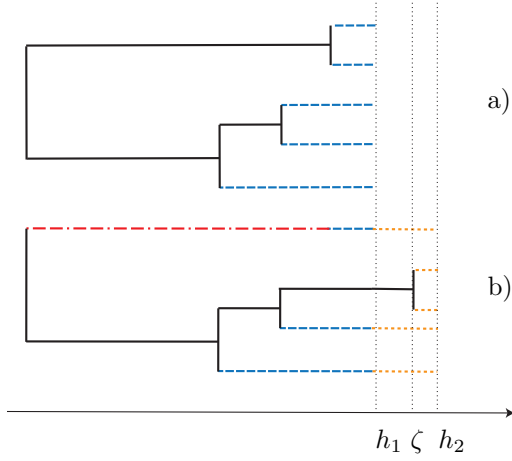


Figure 2: In Figure a) the external branches of the coalescent at time h_1 are marked in (dashed) blue. At time ζ a reproduction event happens and the tree changes. The external branches at time h_2 are shown in Figure b) with the following code: the (dashed) blue pieces make up the part of the external length from time h_1 that is still external at time h_2 , the red (dash and dot) piece is a branch that was internal at time h_1 and became free at time ζ , while the orange (dotted) pieces represent the external length I_{h_1, h_2}^n gathered in the time interval (h_1, h_2)

A key role in the analysis of the external length process is played by the main result of [5] which states that for any $h \in \mathbb{R}$ and any $r \in \mathbb{N}$, as $n \rightarrow \infty$

$$\sqrt{\frac{n}{4 \log n}} \left(\mathcal{L}_h^{n,1} - \eta_1, \dots, \mathcal{L}_h^{n,r} - \eta_r \right) \xrightarrow{d} N(0, I_r),$$

where I_r denotes the $r \times r$ -identity matrix and

$$\eta_i = \mathbb{E}(\mathcal{L}_h^{n,i}) = \frac{2}{i} \quad \text{for every } i \geq 1. \quad (2)$$

Let us label the individuals alive at time 0 by $1, \dots, n$ and denote by \mathcal{P}_n the power set of $\{1, \dots, n\}$. For $A \in \mathcal{P}_n \setminus \{\emptyset, \{1, \dots, n\}\}$ denote

$L_A :=$ the length of the branch supporting the leaves with labels in A in the coalescent at time 0,

with the convention $L_A = 0$ if there is no such branch in the coalescent and

$\mathcal{F}_A := \{\text{the number of descendants at time } h \text{ of the individuals with labels in } A \text{ is equal to } 1\}.$

Observe that with these notations

$$\mathbf{1}_{\mathcal{F}_A} \cdot L_A = \mathbf{1}_{\{\text{there is a branch supporting the leaves with labels in } A \text{ at time 0 which is free at time } h\}} \cdot L_A$$

and thus the i -th summand on the right-hand side of (1) can be rewritten as

$$\Lambda^{n,i} \mathcal{L}_0^{n,i} = \sum_{\substack{A \in \mathcal{P}_n \\ |A|=i}} \mathbf{1}_{\mathcal{F}_A} \cdot L_A.$$

For any fixed $r \in \mathbb{N}$ we can write (1) as

$$\mathcal{L}_h^n = \sum_{\substack{A \in \mathcal{P}_n \\ 1 \leq |A| \leq r}} \mathbf{1}_{\mathcal{F}_A} \cdot L_A + R_h^{n,r} + I_{0,h}^n, \quad (3)$$

where

$$R_h^{n,r} := \sum_{\substack{A \in \mathcal{P}_n \\ r < |A| \leq n}} \mathbf{1}_{\mathcal{F}_A} \cdot L_A \quad (4)$$

records the contribution of the lengths of the branches of orders larger than r to the external length at time h .

The proof of Theorem 1 will show that for large r it is only the first term on the right-hand side of (3) that makes a non-negligible contribution to the external length at time h . Note moreover that this sum is obtained by a random thinning (by means of the indicators $\mathbf{1}_{\mathcal{F}_A}$) of the lengths $\mathcal{L}_0^{n,i}$ of order at most r . In Section 2 we prove a general result, which assures that if a sum of (possibly dependent) random variables converging in distribution to a normal-distributed random variable, is endowed with independent random coefficients, then the sum obtained through this procedure continues to be asymptotically Gaussian. The result of [5] provides the asymptotic normality assumption on the initial sums $\mathcal{L}_0^{n,i}$.

The second key idea of the paper is that the *freeing*-mechanism described above can be coupled in a natural way with a birth and death process started with $|A|$ individuals at time 0. This allows to replace the original thinning procedure by another one which provides independent coefficients. Moreover, the coupling is also used in the investigation of the contributions that the branches carrying a large number of leaves make to the evolving external length (in Section 5). The coupling is presented in Section 3.

The third building block of the proof of our main result is a new description of the coalescent tree structure (also called the tree topology). This is obtained by neglecting the exponential lengths of the branches. Then the whole tree structure is captured by the random variables K_A and J_A , $A \in \mathcal{P}_n \setminus \{\emptyset, \{1, \dots, n\}\}$, where K_A is the level in the coalescent at which the branch carrying the leaves with labels in the set A is formed and J_A is the level at which this branch ends (with the convention that $K_A = J_A = \infty$ if no such branch exists in the coalescent). We show that these quantities (and not the exponential times) play the central role in the development of the external length. We examine them in detail in Section 4.

Building on this preparatory work and on Fu's estimates [13] we deal in Section 5 with the expectation and variance of the length contributed to the evolving external length by the large families from the coalescent at time 0.

The proof of Theorem 1 is then completed in Section 6. Putting the building blocks described above together one more idea is needed for the proof, namely it is necessary to distinguish between the contributions to the evolving external length coming from the possible freeing of the internal lengths that are in small, medium-sized and respectively large depth of the tree. It turns out that for the contributions of the first and the latter groups of branches one can easily replace the freeing procedures by procedures dictated by independent birth and death processes (for each branch separately). In the case of the second group of internal branches (defined as the branches crossing a given fixed level in the tree), it holds that the sets of leaves carried by these branches are all disjoint. This property allows us to apply the coupling from Section 3 for all branches in the group simultaneously in order to obtain independent coefficients for the thinning. By the replacement of the freeing-mechanism through independent birth and death processes we also obtain the formula for the covariance of the limiting Gaussian process: $\left(\frac{2}{2+h}\right)^2$ is simply the probability that a critical birth and death process starting with one individual at time 0 has one individual at time h .

Notation. We recall the Vinogradov notation:

$$a_n \ll b_n \text{ if there exists a finite constant } c \text{ independent of } n \\ \text{(but possibly dependent of } r \text{ and } h) \text{ such that } |a_n| \leq cb_n.$$

We use this notation instead of the O-notation since for voluminous formulas it is easier to read.

Moreover we denote throughout by c a finite constant whose value is not important and may change from line to line.

2 Asymptotic normality of sums with random coefficients

The following Proposition contains a general statement of the form: If the sum of not necessarily independent random variables is asymptotically normal then this property persists if the summands are supplied with independent random coefficients.

Proposition 1. *Let $r \in \mathbb{N}$ be fixed and let $k_1(n), \dots, k_r(n) \in \mathbb{N}$ be such that $\lim_{n \rightarrow \infty} k_i(n) = \infty$ for each $1 \leq i \leq r$. Let $Y_{i,j}, 1 \leq i \leq r, 1 \leq j \leq k_i(n)$ be random variables such that in \mathbb{R}^r*

$$\left(\sum_{j=1}^{k_i(n)} Y_{i,j} \right)_{i=1, \dots, r} \xrightarrow{d} N(0, I_r).$$

Also let $U_{i,j}, 1 \leq i \leq r, 1 \leq j \leq k_i(n)$ be independent of one another and independent of the collection $\{Y_{i,j}\}_{\substack{1 \leq i \leq r \\ 1 \leq j \leq k_i(n)}}$, with the property that for fixed index i the variables $U_{i,j}, 1 \leq j \leq k_i(n)$ are identically distributed with second moment m_i . If the following two conditions hold: for each $1 \leq i \leq r$

$$\sum_{j=1}^{k_i(n)} Y_{i,j}^2 \xrightarrow{\mathbb{P}} 1 \quad \text{and} \quad \max_{1 \leq j \leq k_i(n)} |Y_{i,j}| \xrightarrow{\mathbb{P}} 0$$

then

$$\sum_{i=1}^r \sum_{j=1}^{k_i(n)} U_{i,j} Y_{i,j} \xrightarrow{d} N\left(0, \sum_{i=1}^r m_i\right).$$

Proof. The characteristic function of the sum $\sum_{i=1}^r \sum_{j=1}^{k_i(n)} U_{i,j} Y_{i,j}$ is

$$\varphi^n(\lambda) = \mathbb{E}\left(\mathbb{E}\left(\exp\left(i\lambda \sum_{i=1}^r \sum_{j=1}^{k_i(n)} U_{i,j} Y_{i,j}\right) \mid Y_{i,j}, i \leq r, j \leq k_i(n)\right)\right) = \mathbb{E}\left(\prod_{i=1}^r \prod_{j=1}^{k_i(n)} \varphi_{U_{i,j}}(\lambda Y_{i,j})\right). \quad (5)$$

Using the fact that for complex numbers $\{a_k\}_{1 \leq k \leq n}$ and $\{b_k\}_{1 \leq k \leq n}$ with $|a_k| \leq 1$ and $|b_k| \leq 1$ for all k it holds that

$$\left| \prod_{k=1}^n a_k - \prod_{k=1}^n b_k \right| \leq \sum_{k=1}^n |a_k - b_k|$$

we obtain by denoting the expectation and variance of $U_{i,j}$ by μ_i and σ_i^2 respectively that

$$\left| \prod_{i,j} \varphi_{U_{i,j}}(\lambda Y_{i,j}) - \prod_{i,j} \exp\left(i\mu_i \lambda Y_{i,j} - \frac{\sigma_i^2}{2} \lambda^2 Y_{i,j}^2\right) \right| \leq \sum_{i,j} \left| \varphi_{U_{i,j}}(\lambda Y_{i,j}) - \exp\left(i\mu_i \lambda Y_{i,j} - \frac{\sigma_i^2}{2} \lambda^2 Y_{i,j}^2\right) \right|$$

and by using a Taylor expansion that

$$\begin{aligned} \left| \prod_{i,j} \varphi_{U_{i,j}}(\lambda Y_{i,j}) - \prod_{i,j} \exp\left(i\mu_i \lambda Y_{i,j} - \frac{\sigma_i^2}{2} \lambda^2 Y_{i,j}^2\right) \right| &\leq \sum_{i,j} \left| 1 + i\mu_i \lambda Y_{i,j} - \frac{1}{2} m_i \lambda^2 Y_{i,j}^2 + o(|Y_{i,j}|^2) \right. \\ &\quad \left. - \left(1 + i\mu_i \lambda Y_{i,j} - \frac{\sigma_i^2}{2} \lambda^2 Y_{i,j}^2 - \frac{\mu_i^2}{2} \lambda^2 Y_{i,j}^2 + O(|Y_{i,j}|^3) \right) \right| \\ &= \sum_{i,j} o(|Y_{i,j}|^2). \end{aligned}$$

By assumption $\sum_{i,j} o(|Y_{i,j}|^2)$ converges in probability to 0 as $n \rightarrow \infty$. It follows by dominated convergence that

$$\mathbb{E}\left(\left| \prod_{i,j} \varphi_{U_{i,j}}(\lambda Y_{i,j}) - \prod_{i,j} \exp\left(i\mu_i \lambda Y_{i,j} - \frac{\sigma_i^2}{2} \lambda^2 Y_{i,j}^2\right) \right|\right) \xrightarrow{n \rightarrow \infty} 0. \quad (6)$$

From the assumptions on the collection of random variables $\{Y_{i,j}\}_{\substack{1 \leq i \leq r \\ 1 \leq j \leq k_i(n)}}$ it follows that $(\sum_j Y_{i,j}, \sum_j Y_{i,j}^2)_i$ converges in distribution and we obtain from (5) and (6) that

$$\begin{aligned} \lim_{n \rightarrow \infty} \varphi^n(\lambda) &= \lim_{n \rightarrow \infty} \mathbb{E} \left(\prod_{i=1}^r \prod_{j=1}^{k_i(n)} \exp(i\mu_i \lambda Y_{i,j} - \frac{\sigma_i^2}{2} \lambda^2 Y_{i,j}^2) \right) \\ &= \lim_{n \rightarrow \infty} \mathbb{E} \left(\exp(i\lambda \sum_{i=1}^r \mu_i \sum_{j=1}^{k_i(n)} Y_{i,j} - \sum_{i=1}^r \lambda^2 \frac{\sigma_i^2}{2} \sum_{j=1}^{k_i(n)} Y_{i,j}^2) \right) \\ &= e^{-\frac{\lambda^2}{2} \sum_{i=1}^r \mu_i^2} \cdot e^{-\frac{\lambda^2}{2} \sum_{i=1}^r \sigma_i^2} \\ &= e^{-\frac{\lambda^2}{2} \sum_{i=1}^r m_i}. \end{aligned}$$

Since the right hand side is the characteristic function of the normal distribution with expectation 0 and variance $\sum_{i=1}^r m_i$ the proof of the proposition is finished. \square

3 Coupling a critical branching process to the Moran model

In the time interval $(0, h)$ on the generations time-scale the individuals in the population are subject to binary reproduction events and death events according to the Moran dynamics with rate $\frac{1}{n}$ per pair. In this section we give the description of the evolution of the number of descendants of the individuals alive at time 0 in terms of independent linear birth and death processes.

We call the descendants of the individual with label $j \in \{1, \dots, n\}$ at time 0, the $\{j\}$ -family and we are interested in the size of the family (number of individuals at each time point). We denote by

$$Z^j := (Z_h^j)_{h \geq 0}$$

the process recording the size of the $\{j\}$ -family as time evolves. Then $Z_0^j = 1$ and according to the Moran dynamics of the population, the process Z^j is a continuous time Markov process with jump rates

$$r_{k,k'} = \begin{cases} \frac{1}{2}k(1 - \frac{k}{n}) & \text{if } k' = k + 1 \\ \frac{1}{2}k(1 - \frac{k}{n}) & \text{if } k' = k - 1. \end{cases}$$

Indeed, it is only the reproduction events that happen between a member of the $\{j\}$ -family and a non- $\{j\}$ -family member, that affect the size Z^j of the $\{j\}$ -family. If Z^j is in state k there are $k(n - k)$ such pairs and the reproduction events arise at rate $\frac{1}{n}$. The family size may either decrease by one or increase by one at the time of such a reproduction event, depending on whether the individual from the $\{j\}$ -family dies or reproduces. The reproduction events between two $\{j\}$ -family members or between two non- $\{j\}$ -family members leave Z^j unchanged.

Let us now fix $A \in \mathcal{P}_n$ and let $i = |A|$, $1 \leq i \leq n - 1$. Denote by

$$Z^A = (Z_h^A)_{h \geq 0}$$

the process which records the size of the A -family as time evolves,

$$Z^A := \sum_{j \in A} Z^j.$$

To the processes Z^j , $j \in A$, we now couple jointly linear birth and death processes $B^j = B^{j,A}$. To this aim, let us first observe that the reproduction events that affect the size Z^j of the $\{j\}$ -family, $j \in A$, can be grouped as follows: on one hand there are the events between an individual from the $\{j\}$ -family and an individual from the $(\{1, \dots, n\} \setminus A)$ -family (we call these events α -events) and on the other hand there are the

reproduction events between an individual from the $\{j\}$ -family and an individual from the $(A \setminus \{j\})$ -family (we call these events β -events). If the process Z^j is in state k and the process Z^A is in state l then the α -events occur at rate $k(n-l)/n$, whereas the β -events occur at rate $k(l-k)/n$.

We construct the coupled processes B^j starting from the processes Z^j , $j \in A$, in two steps. The first step is to leave aside the influence of the β -events on the size of the $\{j\}$ -family. This gives rise to a process Y^j which, when Z^A is in state l , jumps at rate

$$\tilde{r}_{k,k'} = \begin{cases} \frac{1}{2}k(1 - \frac{l}{n}) & \text{if } k' = k + 1 \\ \frac{1}{2}k(1 - \frac{l}{n}) & \text{if } k' = k - 1. \end{cases}$$

The second step of the construction is to perform a random time change $\theta_h = \theta_h^A$ on the process Y^j depending on the state of the process Z^A . Note that n is an absorbing state for the process Z^A and denote by

$$\tau = \tau^A := \inf\{u \geq 0 : Z_u^A = n\}$$

the hitting time of the process Z^A at state n .

Let

$$\theta_h = \theta_h^A := \int_0^h \left(1 - \frac{Z_u^A}{n}\right) du. \quad (7)$$

Observe that θ_h as a function of h is continuous and strictly increasing for $h \leq \tau$. With this notation we define B_θ^j for $\theta \leq \theta_\tau$ implicitly by

$$B_{\theta_h}^j = B_{\theta_h}^{j,A} := Y_h^j, \quad h \leq \tau. \quad (8)$$

Note that $\theta_h < h$ for all $0 < h \leq \tau$, thus B^j is a randomly speeded-up version of the process Y^j . The process B^j is (using results of [12] Section 6.1) a birth and death process with birth and death rates both equal to $\frac{1}{2}k$ when the process is in state k . In order to fully define the processes B^j , $j \in A$, we let them evolve from the states $B_{\theta_\tau}^j$ according to these rates independently (of each other and of everything else).

Observe from the construction given above for the processes B^j , $j \in A$, that the β -events (which are eliminated for each j in order to obtain the process Y^j from Z^j) do not affect the size Z^A of the A -family since they take place between two individuals of this family (one individual reproduces and another dies and the size of the A -family remains constant). Thus

$$Z^A = \sum_{j \in A} Z^j = \sum_{j \in A} Y^j =: Y^A \quad (9)$$

and as a consequence by (8) for all $h \leq \tau$

$$B_{\theta_h}^A = Z_h^A, \quad (10)$$

where $B^A := \sum_{j \in A} B^j$.

Define the σ -algebra

$$\mathcal{F}_h := \sigma(B_s^j, s \leq h, j \in A)$$

and observe that

$$\theta_\tau = \inf\{u \geq 0 : B_u^A = n\}.$$

Thus, θ_τ is a stopping time with respect to the filtration $\mathcal{F} := (\mathcal{F}_h)_{h \geq 0}$. For $h \leq \tau$ the relation (7) can also be expressed as

$$h = \int_0^{\theta_h} \frac{1}{1 - \frac{B_u^A}{n}} du \quad (11)$$

and thus for any $h \geq 0$ it holds

$$\{\theta_h < s\} = \{\theta_\tau < s\} \cup \left\{ \theta_\tau \geq s, \int_0^s \frac{1}{1 - \frac{B_u^A}{n}} du > h \right\} \in \mathcal{F}_s.$$

Therefore θ_h is also a stopping time with respect to the filtration \mathcal{F} .

Lemma 1. *The processes B^j , $j \in A$, defined above are independent birth and death processes and also independent of the coalescent tree at time 0.*

Proof. Without loss of generality we consider $A = \{1, \dots, i\}$. Then from the construction given above, since we let the β -events aside, $Y = (Y^1, \dots, Y^i)$ is a Markov process which jumps, when $Z^A = Y^A$ is in state l , at rate

$$\lambda_{(k_1, \dots, k_j, \dots, k_i), (k_1, \dots, k'_j, \dots, k_i)} = \begin{cases} \frac{1}{2}k_j(1 - \frac{l}{n}) & \text{if } k'_j = k_j + 1 \\ \frac{1}{2}k_j(1 - \frac{l}{n}) & \text{if } k'_j = k_j - 1. \end{cases}$$

By performing the random time change from above on the process Y one then obtains the Markov process (B^1, \dots, B^i) with jumps rates

$$\tilde{\lambda}_{(k_1, \dots, k_j, \dots, k_i), (k_1, \dots, k'_j, \dots, k_i)} = \begin{cases} \frac{1}{2}k_j & \text{if } k'_j = k_j + 1 \\ \frac{1}{2}k_j & \text{if } k'_j = k_j - 1. \end{cases}$$

This implies the first assertion. The independence of the coalescent tree at time 0 follows from the independence properties of the underlying Moran model. \square

Before further analysing the relation between the processes B^j and Z^j , $j \in A$, we compute a few basic quantities connected to B^j , that we will use in the sequel. Let

$$F(s, h) = \sum_{k=0}^{\infty} s^k \mathbb{P}(B_h^j = k)$$

denote the generating function of the process B^j . Then (see [1], Chapter III) F is the unique solution of the backward equation

$$\frac{\partial}{\partial h} F(s, h) = \frac{1 + F^2(s, h)}{2} - F(s, h) = \frac{1}{2}(1 - F(s, h))^2$$

with boundary condition $F(s, 0) = s$. Integrating the equation gives

$$F(s, h) = 1 - \frac{1 - s}{1 + \frac{h}{2} - \frac{h}{2}s}, \quad (12)$$

which can be rewritten as

$$F(s, h) = 1 - \frac{1 - s}{1 + \frac{h}{2}} \cdot \sum_{j=0}^{\infty} \left(\frac{h}{h+2} \cdot s \right)^j.$$

This gives

$$\mathbb{P}(B_h^j = 0) = \frac{h}{h+2} \quad \text{and} \quad \mathbb{P}(B_h^j = 1) = \left(\frac{2}{h+2} \right)^2. \quad (13)$$

The process B^A has the generating function F^i and it holds that

$$\mathbb{P}(B_h^A = 1) = i \cdot \left(\frac{h}{h+2} \right)^{i-1} \left(\frac{2}{h+2} \right)^2. \quad (14)$$

Observe that for fixed h the generating function F and all its derivatives exist for all $s < \frac{h+2}{h}$ and therefore all moments of B_h^A exist and are finite functions of h . For $d \in \mathbb{N}$

$$\mathbb{E}\left((B_h^A)^d\right) = \mathbb{E}\left(\left(\sum_{j \in A} B_h^j\right)^d\right).$$

Expanding the sum it follows that

$$\mathbb{E}\left((B_h^A)^d\right) \leq c_h \cdot i^d, \quad (15)$$

for a constant c_h depending on h .

In the rest of this section we focus on the relationship between the processes Z^j and B^j . For $A' \subset A$, $|A'| = i'$ denote similarly as before $Z^{A'} = \sum_{j \in A'} Z^j$, $Y^{A'} = Y^{A', A} = \sum_{j \in A'} Y^j$ and $B^{A'} = B^{A', A} = \sum_{j \in A'} B^j$. Note that contrarily to $Z^{A'}$, the distributions of the processes $Y^{A'}$ and $B^{A'}$ depend on the set A which we fixed in advance.

Lemma 2. *For $A' \subset A$, $|A'| = i' \leq i$ the following bounds hold*

$$\mathbb{P}(Y_h^{A'} = 1) \leq 2i' \cdot \left(\frac{h}{h+2}\right)^{i'-1} \quad (16)$$

and

$$\mathbb{P}(B_h^{A'} = 1, Z_h^{A'} \neq 1) + \mathbb{P}(Z_h^{A'} = 1, B_h^{A'} \neq 1) \leq c \cdot \frac{i}{n} \cdot \gamma^{i'} \quad (17)$$

for some $\gamma < 1$ and $c < \infty$ depending only on h .

When $A' = A$ the inequality (16) becomes

$$\mathbb{P}(Z_h^A = 1) \leq 2i \cdot \left(\frac{h}{h+2}\right)^{i-1}. \quad (18)$$

Proof. We write for short $Z^{A'} = Z'$, $Y^{A'} = Y'$ and $B^{A'} = B'$ and define the random variables

$$N(Y', h) := \text{the number of the individuals in } A' \text{ that have 0 descendants at time } h, \\ \text{when evolving according to } Y'$$

and $N(B', h)$ analogously.

Observe from the definition of $Y^j, j \in A$, that from the time τ on, the state of the process Y^j remains unchanged, that is

$$Y_h^j = Y_\tau^j \quad \text{for all } h \geq \tau.$$

Thus, by (8) it follows that for any $h \geq 0$

$$Y_h^j = B_{\theta_h \wedge \theta_\tau}^j. \quad (19)$$

Since $Y'_0 = i'$ and $\theta_h \wedge \theta_\tau < h$ we obtain that for any $h \geq 0$

$$\{Y'_h = 1\} \subset \{N(Y', h) = i' - 1\} = \{N(B', \theta_h \wedge \theta_\tau) = i' - 1\} \subset \{N(B', h) \geq i' - 1\}.$$

Using (13) we thus obtain that

$$\mathbb{P}(Y'_h = 1) \leq i' \cdot \left(\frac{h}{h+2}\right)^{i'-1} + \left(\frac{h}{h+2}\right)^{i'} \leq 2i' \cdot \left(\frac{h}{h+2}\right)^{i'-1},$$

which gives the first claim of the lemma.

As to the second claim observe that the β -events between a member of the $\{A'\}$ -family and a member of the $(A \setminus A')$ -family are the β -events that lead to changes in the size Z' of the A' -family. We call these events $\beta^{A'}$ -events or for short β' -events. These occur at rate $k(l-k)/n$ when the processes Z' and Z^A are in the states k and l respectively.

Let

$$\rho = \rho^{A,A'} := \inf\{u \geq 0 : \text{a } \beta'\text{-event takes place at time } u\}$$

and

$$E = E_{\beta',h} := \{\rho < h\}.$$

Note that on E^c it holds that

$$Z'_s = Y'_s \quad \text{for all } s \in [0, h). \quad (20)$$

Then

$$\begin{aligned} & \mathbb{P}(B'_h = 1, Z'_h \neq 1) + \mathbb{P}(Z'_h = 1, B'_h \neq 1) \\ & \leq \mathbb{P}(B'_h = 1, E) + \mathbb{P}(B'_h = 1, Y'_h \neq 1) + \mathbb{P}(Z'_h = 1, E) + \mathbb{P}(Y'_h = 1, B'_h \neq 1). \end{aligned} \quad (21)$$

Observe that since the β' -events do not affect the evolution of Y' and Y^A , in view of (8), (9), (11) and (20), given Y' and Y^A , the first β' -event occurs at rate $\frac{Y'_s(Y_s^A - Y'_s)}{n} \cdot \exp\left(-\int_0^s \frac{Y'_u(Y_u^A - Y'_u)}{n} du\right) ds$. Thus for the first term on the right-hand side of (21) we obtain

$$\begin{aligned} \mathbb{P}(B'_h = 1, E) &= \mathbb{E}\left(\mathbb{P}(E \mid \mathcal{F}_h); B'_h = 1\right) \\ &\leq \mathbb{E}\left(\int_0^h \frac{Y'_s(Y_s^A - Y'_s)}{n} ds; B'_h = 1\right) \\ &\leq \frac{1}{n} \mathbb{E}\left(\int_0^h Y'_s Y_s^A ds; B'_h = 1\right) \end{aligned} \quad (22)$$

and by Fubini's theorem and Hölder's inequality

$$\mathbb{P}(B'_h = 1, E) \leq \frac{1}{n} \int_0^h \mathbb{E}((Y'_s)^3)^{1/3} \cdot \mathbb{E}((Y_s^A)^3)^{1/3} \cdot \mathbb{P}(B'_h = 1)^{1/3} ds.$$

Since the function $x \mapsto x^3$ is convex for $x \geq 0$ and B' is a martingale with respect to the filtration \mathcal{F} , it follows that the process $(B')^3$ is a submartingale. We use (19) together with the fact that θ_s and θ_τ are stopping times and that $\theta_s \leq s$ and obtain for $s \leq h$ that

$$\mathbb{E}((Y'_s)^3) = \mathbb{E}((B'_{\theta_s \wedge \theta_\tau})^3) \leq \mathbb{E}((B'_{\theta_s})^3) \leq \mathbb{E}((B'_s)^3) \leq \mathbb{E}((B'_h)^3). \quad (23)$$

Using (15) we have that for $s \leq h$

$$\mathbb{E}((B'_s)^3) \leq c \cdot i'^3 \quad \text{and} \quad \mathbb{E}((Y'_s)^3) \leq c \cdot i'^3, \quad (24)$$

in particular

$$\mathbb{E}((Y_s^A)^3) \leq c \cdot i^3 \quad (25)$$

for $c < \infty$ depending on h . Now using (14) it follows that

$$\mathbb{P}(B'_h = 1, E) \leq c \cdot \frac{i}{n} \cdot i' \cdot \gamma^{i'}$$

for some $\gamma < 1$ and a finite constant c depending on h . By further enlarging γ we can drop the factor i' and thus obtain that

$$\mathbb{P}(B'_h = 1, E) \leq c \cdot \frac{i}{n} \cdot \gamma^{i'} \quad (26)$$

for $\gamma < 1$ and $c < \infty$ depending on h

For the third term on the right-hand side of (21) observe by using the strong Markov property and the bound for $\mathbb{P}(Z'_h = 1)$ given in (18) for $A' = A$ that

$$\mathbb{P}(Z'_h = 1, E \mid Z'_s, s \leq \rho) = \mathbf{1}_E \cdot \mathbb{P}_{Z'_\rho}(Z'_{h-\rho} = 1) \leq \mathbf{1}_E \cdot 2Z'_\rho \left(\frac{h-\rho}{h-\rho+2} \right)^{Z'_\rho-1}$$

and thus for a finite constant c depending on h

$$\mathbb{P}(Z'_h = 1, E) \leq c \cdot \mathbb{E} \left(\mathbf{1}_E \cdot Z'_\rho \left(\frac{h}{h+2} \right)^{Z'_\rho} \right) = c \cdot \mathbb{E} \left(\mathbf{1}_E \cdot \mathbb{E} \left(Z'_\rho \left(\frac{h}{h+2} \right)^{Z'_\rho} \mid Z'_s, s < \rho \right) \right).$$

Since $Z'_s = Y'_s$ for all times $s < \rho$ and at the time of a β' -event the process Z' jumps by either +1 or -1 with equal probability and independently of the past we further deduce that

$$\begin{aligned} \mathbb{P}(Z'_h = 1, E) &\leq c \cdot \mathbb{E} \left(\mathbf{1}_E \cdot \left(\frac{1}{2}(Y'_\rho + 1) \left(\frac{h}{h+2} \right)^{Y'_\rho+1} + \frac{1}{2}(Y'_\rho - 1) \left(\frac{h}{h+2} \right)^{Y'_\rho-1} \right) \right) \\ &\leq c \cdot \mathbb{E} \left(\mathbf{1}_E \cdot (Y'_\rho + 1) \left(\frac{h}{h+2} \right)^{Y'_\rho} \right) \end{aligned}$$

and similarly as in (22)

$$\begin{aligned} \mathbb{P}(Z'_h = 1, E) &\leq c \cdot \mathbb{E} \left(\int_0^h (Y'_s + 1) \left(\frac{h}{h+2} \right)^{Y'_s} \cdot \frac{1}{n} \cdot Y'_s (Y_s^A - Y'_s) ds \right) \\ &\leq c \cdot \frac{1}{n} \cdot \mathbb{E} \left(\int_0^h Y'^2 Y_s^A \cdot \left(\frac{h}{h+2} \right)^{Y'_s} ds \right). \end{aligned}$$

Now by Fubini's theorem and Hölder's inequality we obtain that

$$\mathbb{P}(Z'_h = 1, E) \leq c \cdot \frac{1}{n} \int_0^h \mathbb{E}((Y'_s)^6)^{1/3} \cdot \mathbb{E}((Y_s^A)^3)^{1/3} \cdot \mathbb{E} \left(\left(\frac{h}{h+2} \right)^{3Y'_s} \right)^{1/3} ds. \quad (27)$$

Since the function $x \mapsto \left(\frac{h}{h+2} \right)^{3x}$ is convex and the process B' is a martingale, the process $\left(\frac{h}{h+2} \right)^{3B'}$ is a submartingale. We thus get by (19) that for $s \leq h$

$$\mathbb{E} \left(\left(\frac{h}{h+2} \right)^{3Y'_s} \right) = \mathbb{E} \left(\left(\frac{h}{h+2} \right)^{3B'_{\theta_s \wedge \theta_\tau}} \right) \leq \mathbb{E} \left(\left(\frac{h}{h+2} \right)^{3B'_h} \right)$$

and moreover

$$\mathbb{E} \left(\left(\frac{h}{h+2} \right)^{3Y'_s} \right) \leq F \left(\left(\frac{h}{h+2} \right)^3, h \right)^{i'} \quad (28)$$

where the generating function F is given in (12). Now using the fact that $F(u, h) < 1$ for $u < 1$ we obtain by plugging this together with (24) and (25) in (27) that

$$\mathbb{P}(Z'_h = 1, E) \leq c \cdot \frac{i}{n} \cdot \gamma^{i'}. \quad (29)$$

for $\gamma < 1$ and $c < \infty$ depending on h .

Let us now turn to the second term on the right-hand side of (21). By (19) it holds that

$$\mathbb{P}(B'_h = 1, Y'_h \neq 1) = \mathbb{P}(B'_h = 1, B'_{\theta_h \wedge \theta_\tau} \neq 1) = \mathbb{E}(\mathbb{P}(B'_h = 1 \mid \mathcal{F}_{\theta_h \wedge \theta_\tau}); B'_{\theta_h \wedge \theta_\tau} \neq 1)$$

and by (14)

$$\mathbb{P}(B'_h = 1, Y'_h \neq 1) \leq \mathbb{E} \left(B'_{\theta_h \wedge \theta_\tau} \left(\frac{h - \theta_h \wedge \theta_\tau}{h - \theta_h \wedge \theta_\tau + 2} \right)^{B'_{\theta_h \wedge \theta_\tau} - 1}; B'_{\theta_h \wedge \theta_\tau} \geq 2 \right) \quad (30)$$

$$\leq \mathbb{E}\left((h - \theta_h \wedge \theta_\tau) \cdot B'_{\theta_h \wedge \theta_\tau} \cdot \left(\frac{h}{h+2}\right)^{B'_{\theta_h \wedge \theta_\tau} - 2}\right).$$

By Hölder's inequality we obtain that

$$\mathbb{P}(B'_h = 1, Y'_h \neq 1) \leq \mathbb{E}((h - \theta_h \wedge \theta_\tau)^3)^{1/3} \cdot \mathbb{E}((B'_{\theta_h \wedge \theta_\tau})^3)^{1/3} \cdot \mathbb{E}\left(\left(\frac{h}{h+2}\right)^{3(B'_{\theta_h \wedge \theta_\tau} - 2)}\right)^{1/3}. \quad (31)$$

Now using that $Z^A = Y^A$ and the fact that the state of the process Y^j , $j \in A$, remains unchanged from time τ on, we get that

$$h - \theta_h \wedge \theta_\tau \leq h - \theta_h + h \cdot \mathbf{1}_{\{\theta_h > \theta_\tau\}} \leq h - \theta_h + h \cdot \frac{Y_h^A}{n}.$$

By (7) and (10)

$$h - \theta_h \wedge \theta_\tau \leq \int_0^h \frac{B_{\theta_s}^A}{n} ds + h \cdot \frac{Y_h^A}{n} = \int_0^h \frac{B_{\theta_s}^A + Y_h^A}{n} ds.$$

Using Jensen's inequality and Fubini's theorem it follows that

$$\mathbb{E}\left((h - \theta_h \wedge \theta_\tau)^3\right) \leq \mathbb{E}\left(h^3 \int_0^h \left(\frac{B_{\theta_s}^A + Y_h^A}{n}\right)^3 \frac{ds}{h}\right) = \frac{h^2}{n^3} \int_0^h \mathbb{E}\left((B_{\theta_s}^A + Y_h^A)^3\right) ds.$$

By (23) and (24) (for $A' = A$) together with (25) we get that for $c < \infty$ depending on h

$$\mathbb{E}\left((h - \theta_h \wedge \theta_\tau)^3\right) \leq c \cdot \frac{i^3}{n^3}. \quad (32)$$

By the same arguments used to obtain (28) and using again (23) and (24) in (31) we obtain that

$$\mathbb{P}(B'_h = 1, Y'_h \neq 1) \leq c \cdot \frac{1}{n} \cdot i \cdot \gamma^{i'} \quad (33)$$

for $\gamma < 1$ and $c < \infty$ depending on h .

For the last term on the left-hand side of (21) we deduce that

$$\begin{aligned} \mathbb{P}(Y'_h = 1, B'_h \neq 1) &= \mathbb{P}(B'_{\theta_h \wedge \theta_\tau} = 1, \exists \text{ at least one birth or death event in } (\theta_h \wedge \theta_\tau, h)) \\ &= \mathbb{E}\left(\mathbb{P}(\exists \text{ at least one birth or death event in } (\theta_h \wedge \theta_\tau, h) \mid \mathcal{F}_{\theta_h \wedge \theta_\tau}); B'_{\theta_h \wedge \theta_\tau} = 1\right) \\ &\leq \mathbb{E}\left(1 - e^{-(h - \theta_h \wedge \theta_\tau)}; B'_{\theta_h \wedge \theta_\tau} = 1\right) \\ &\leq \mathbb{E}(h - \theta_h \wedge \theta_\tau; Y'_h = 1). \end{aligned}$$

Using Hölder's inequality together with (32) and (16) we obtain that

$$\mathbb{P}(Y'_h = 1, B'_h \neq 1) \leq c \cdot \frac{1}{n} \cdot i \cdot \gamma^{i'} \quad (34)$$

for some $\gamma < 1$ and $c < \infty$ depending on h .

This together with (26), (29) and (33) gives the second claim of the lemma. \square

Let the set A continue to be fixed, consider $A' \subset A$ with $|A'| = i'$ and denote $A'' := A \setminus A'$ and $i'' := i - i' = |A''|$. As before consider the associated processes $Z'' := Z^{A''}$, $Y'' := Y^{A''}$ and $B'' := B^{A''}$.

Lemma 3. *The following bound holds*

$$\begin{aligned} & \mathbb{P}(B_h^{A'} = 1, Z_h^{A'} \neq 1, Z_h^{A''} = 1) + \mathbb{P}(Z_h^{A'} = 1, B_h^{A'} \neq 1, Z_h^{A''} = 1) \\ & + \mathbb{P}(B_h^{A'} = 1, Z_h^{A'} \neq 1, B_h^{A''} = 1) + \mathbb{P}(Z_h^{A'} = 1, B_h^{A'} \neq 1, B_h^{A''} = 1) \\ & \leq c \cdot \frac{1}{n} \cdot \gamma^i. \end{aligned} \quad (35)$$

for some $\gamma < 1$ and $c < \infty$ depending on h .

Proof. We use similar arguments as in the proof of Lemma 2. Note from definition of the $\beta^{A'}$ -events (which are actually also the $\beta^{A''}$ -events) that these events involve an individual from the A' - and one from the A'' -family and that they lead to changes in the sizes of the two families.

It holds that

$$\{B'_h = 1, Z'_h \neq 1, Z''_h = 1\} \subset \{B'_h = 1, Z''_h = 1, E\} \cup \{B'_h = 1, Y''_h = 1, Y'_h \neq B'_h\}$$

and similarly for the other terms in (35). Therefore the left-hand side of (35) is bounded from above by

$$\begin{aligned} & \mathbb{P}(B'_h = 1, Z''_h = 1, E) + \mathbb{P}(B'_h = 1, Y''_h = 1, Y'_h \neq B'_h) \\ & + \mathbb{P}(Z'_h = 1, Z''_h = 1, E) + \mathbb{P}(Y'_h = 1, Y''_h = 1, Y'_h \neq B'_h) \\ & + \mathbb{P}(B'_h = 1, B''_h = 1, E) + \mathbb{P}(B'_h = 1, B''_h = 1, Y'_h \neq B'_h) \\ & + \mathbb{P}(Z'_h = 1, B''_h = 1, E) + \mathbb{P}(Y'_h = 1, B''_h = 1, Y'_h \neq B'_h). \end{aligned} \quad (36)$$

We start by estimating the first term on the right hand side. Using Hölder's inequality, (26) and (29) we obtain for some $\gamma < 1$ and $c < \infty$ depending on h that

$$\mathbb{P}(B'_h = 1, Z''_h = 1, E) \leq \mathbb{P}(B'_h = 1, E)^{1/2} \cdot \mathbb{P}(Z''_h = 1, E)^{1/2} \leq c \cdot \frac{1}{n} \cdot i \cdot \gamma^{i' + i''}.$$

Since $i = i' + i''$ we can further enlarge γ to incorporate i and therefore obtain

$$\mathbb{P}(B'_h = 1, Z''_h = 1, E) \leq c \cdot \frac{1}{n} \cdot \gamma^i. \quad (37)$$

The same argument works for evaluating the third, fifth and seventh term on the right hand side of (36) and gives the same bound.

In order to evaluate the second term on the right-hand side of (36) we follow the arguments from the proof of Lemma 2 used to show (33) in which we incorporate the event $\{Y''_h = 1\} = \{B''_{\theta_h \wedge \theta_\tau} = 1\}$ which is $\mathcal{F}_{\theta_h \wedge \theta_\tau}$ -measurable. As in (30) we obtain

$$\mathbb{P}(B'_h = 1, Y'_h \neq 1, Y''_h = 1) \leq \mathbb{E}\left((h - \theta_h \wedge \theta_\tau) \cdot B'_{\theta_h \wedge \theta_\tau} \cdot \left(\frac{h}{h+2}\right)^{B'_{\theta_h \wedge \theta_\tau} - 2}; Y''_h = 1\right)$$

which by Hölder's inequality, (16) and using further the arguments that lead to (33) gives

$$\mathbb{P}(B'_h = 1, Y''_h = 1, Y'_h \neq B'_h) \leq c \cdot \frac{1}{n} \cdot \gamma^i \quad (38)$$

for some $\gamma < 1$ and $c < \infty$ depending on h . We used the fact that $i' + i'' = i < n$.

In order to bound the fourth term on the right-hand side of (36) we again a similar argument as in Lemma 2, namely the one used to get (34), where we incorporate the event $\{Y''_h = 1\} = \{B''_{\theta_h \wedge \theta_\tau} = 1\}$ and obtain the desired bound.

For estimating the sixth term on the right-hand side of (36) observe that by Hölder's inequality

$$\begin{aligned}
\mathbb{P}(B'_h = 1, B''_h = 1, Y'_h \neq B'_h) &\leq \mathbb{P}(B'_h = 1, Y'_h \neq B'_h, Y''_h = 1) \\
&\quad + \mathbb{P}(B'_h = 1, Y'_h \neq B'_h, B''_h = 1, Y''_h \neq B''_h) \\
&\leq \mathbb{P}(B'_h = 1, Y'_h \neq B'_h, Y''_h = 1) \\
&\quad + \mathbb{P}(B'_h = 1, Y'_h \neq B'_h)^{1/2} \cdot \mathbb{P}(B''_h = 1, Y''_h \neq B''_h)^{1/2}
\end{aligned}$$

which by means of (33) and (38) can be again bounded from above by $c \cdot \frac{1}{n} \cdot \gamma^i$ for some $\gamma < 1$ and $c < \infty$ depending on h . The same argument works also for the eighth term on the right-hand side of (36). This completes the proof. \square

To finish this section we deduce from the previous results that for $A' \cap A'' = \emptyset$ with $|A'| = i'$ and $|A''| = i''$ the following bound holds

$$|\text{COV}(\mathbf{1}_{\mathcal{F}_{A'}}, \mathbf{1}_{\mathcal{F}_{A''}})| \leq c \cdot \frac{1}{n} \cdot \gamma^{i'+i''}, \quad (39)$$

for some $\gamma < 1$ and $c < \infty$ depending on h .

To see that this is a direct consequence of the previous three lemmas, observe that since the birth-death processes B' and B'' are independent, as resulting from Lemma 1, it holds that

$$\begin{aligned}
\text{COV}(\mathbf{1}_{\mathcal{F}_{A'}}, \mathbf{1}_{\mathcal{F}_{A''}}) &= \text{COV}(\mathbf{1}_{\{Z'_h=1\}}, \mathbf{1}_{\{Z''_h=1\}}) \\
&= \text{COV}(\mathbf{1}_{\{B'_h=1\}}, \mathbf{1}_{\{B''_h=1\}}) \\
&\quad + \text{COV}(\mathbf{1}_{\{B'_h=1\}}, \mathbf{1}_{\{Z''_h=1\}} - \mathbf{1}_{\{B''_h=1\}}) + \text{COV}(\mathbf{1}_{\{Z'_h=1\}} - \mathbf{1}_{\{B'_h=1\}}, \mathbf{1}_{\{Z''_h=1\}}) \\
&= \mathbb{E}(\mathbf{1}_{\{B'_h=1\}}(\mathbf{1}_{\{Z'_h=1\}} - \mathbf{1}_{\{B'_h=1\}})) - \mathbb{P}(B'_h = 1)\mathbb{E}(\mathbf{1}_{\{Z''_h=1\}} - \mathbf{1}_{\{B''_h=1\}}) \\
&\quad + \mathbb{E}(\mathbf{1}_{\{Z''_h=1\}}(\mathbf{1}_{\{Z'_h=1\}} - \mathbf{1}_{\{B'_h=1\}})) - \mathbb{P}(Z''_h = 1)\mathbb{E}(\mathbf{1}_{\{Z'_h=1\}} - \mathbf{1}_{\{B'_h=1\}}) \\
&= \mathbb{P}(B'_h = 1, Z''_h = 1, Z''_h \neq B''_h) - \mathbb{P}(B'_h = 1, B''_h = 1, Z''_h \neq B''_h) \\
&\quad - \mathbb{P}(B'_h = 1)\mathbb{P}(Z''_h = 1, Z''_h \neq B''_h) + \mathbb{P}(B'_h = 1)\mathbb{P}(B''_h = 1, Z''_h \neq B''_h) \\
&\quad + \mathbb{P}(Z''_h = 1, Z'_h = 1, Z'_h \neq B'_h) - \mathbb{P}(Z''_h = 1, B'_h = 1, Z'_h \neq B'_h) \\
&\quad - \mathbb{P}(Z''_h = 1)\mathbb{P}(Z'_h = 1, Z'_h \neq B'_h) + \mathbb{P}(Z''_h = 1)\mathbb{P}(B'_h = 1, Z'_h \neq B'_h).
\end{aligned}$$

Now (39) follows directly by applying Lemma 2, Lemma 3 and (14).

4 The initial and final levels of branches

We start by making some notation. We fix a set $A \in \mathcal{P}_n$, $A \neq \emptyset, \{1, \dots, n\}$, and view the coalescent tree of the population alive at time 0 from the leaves towards the root, that is in the coalescent time direction. Recall that Π_k denotes the state of the coalescent after $n - k$ coalescing events. We denote the levels at which the branch with leaves in A is formed and respectively ends by K_A and J_A , thus

$$K_A := \max\{2 \leq k \leq n : A \in \Pi_k\} \quad \text{and} \quad J_A := \max\{1 \leq j < K_A : A \notin \Pi_j\}.$$

These are the *initial level* and the *final level* of the branch. For a set A of leaves which is not supported by some branch (which means that $A \notin \Pi_k$ for all k) it is convenient to set $K_A = J_A = \infty$. On the event

$\{J_A < \infty\} = \{K_A < \infty\}$ the coalescent contains a subtree with root at level J_A , its last merger at level K_A and its set of leaves equal to A . We denote this subtree by \mathcal{T}_A . The length of the branch supporting the leaves with labels in A is

$$L_A := \sum_{j=J_A+1}^{K_A} X_j, \quad (40)$$

where the $X_j \sim \text{Exp}\left(\binom{j}{2}\right)$ are the inter-coalescence times. For sets A of leaves that are not supported by any branch, we have that $L_A = 0$. Note that the whole tree structure, up to the coalescent times, is captured by the random variables $\{K_A, J_A\}_{A \in \mathcal{P}_n \setminus \{\emptyset, \{1, \dots, n\}\}}$.

The following lemma gives bounds on the distribution weights of the initial and final levels of branches in the coalescent of the population alive at time 0. For the remainder of this section we consider the sets $A, A' \in \mathcal{P}_n$ with $|A| = i$ and $|A'| = i'$ fixed, $1 \leq i, i' < n$.

Lemma 4. *[(i)]*

1. For $1 \leq j < n$ and $1 \leq j' < n$ the following bounds hold:

$$\mathbb{P}(J_A = j) \leq \frac{2j}{(n-1)} \cdot \frac{1}{\binom{n}{i}}$$

and for $A \neq A'$

$$\mathbb{P}(J_A = j, J_{A'} = j') \leq \begin{cases} \frac{4jj'}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i, i', n-(i+i')}} & \text{if } A \cap A' = \emptyset, i + i' < n \\ \frac{4j(i-i')}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i-i', i', n-i}} & \text{if } A' \subset A \text{ and } j' > j. \end{cases}$$

Otherwise $\mathbb{P}(J_A = j, J_{A'} = j') = 0$.

Moreover, if $i + i' < n$ and $A \cap A' = \emptyset$ then for a finite constant c depending on i and i'

$$\mathbb{P}(J_A = j, J_{A'} = j') \leq \mathbb{P}(J = j) \mathbb{P}(J' = j') \left(1 + \frac{c}{n}\right).$$

2. Let $i, i' \geq 2$. Then for $1 \leq j < k < n$ and $1 \leq j' < k' < n$ the following holds:

$$\mathbb{P}(K_A = k) \leq \frac{ik^2}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i}}, \quad \mathbb{P}(K_A = k, J_A = j) \leq \frac{2ij}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i}}.$$

Moreover, if $i + i' < n$ and $A \cap A' = \emptyset$ then for a finite constant c depending on i and i'

$$\mathbb{P}(K_A = k, J_A = j, K_{A'} = k', J_{A'} = j') \leq \mathbb{P}(K_A = k, J_A = j) \mathbb{P}(K_{A'} = k', J_{A'} = j') \left(1 + \frac{c}{n}\right).$$

Remark 1. Note that the bounds of the lemma are not valid in two exceptional cases: if $i = 1$ then $K_A = n$ and if $i + i' = n$ and $A \cap A' = \emptyset$ then $J_A = 1$ implies $J_{A'} = 1$ such that

$$\mathbb{P}(J_A = 1, J_{A'} = 1) = \mathbb{P}(J_A = 1) \leq \frac{2}{(n-1) \cdot \binom{n}{i}}.$$

Proof. For this proof let $J := J_A$, $J' := J_{A'}$, $K := K_A$ and $K' := K_{A'}$.

(i) Note first that if $j > n - i$ then $\mathbb{P}(J = j) = 0$ and the claim holds trivially. Thus let $j \leq n - i$ and denote by E_1 the event that the subtree \mathcal{T}_A has its $i - 1$ mergers at the fixed levels $n > k_1 > \dots > k_{i-1} > j$. To generate this event it is required that from level $n - 1$ to level $j + 1$ the branches from \mathcal{T}_A merge at the levels k_1, \dots, k_{i-1} and that at the remaining levels branches from outside of \mathcal{T}_A merge. In addition, at level j the remaining branch from \mathcal{T}_A has to coalesce with one of the other j branches extant in the coalescent tree.

For the $i - 1$ mergers within \mathcal{T}_A there are $\binom{i}{2} \cdot \binom{i-1}{2} \cdots \binom{2}{2}$ available ways in which the branches can merge with one another, whereas for the other $n - i - j$ mergers there are $\binom{n-i}{2} \cdot \binom{n-i-1}{2} \cdots \binom{j+1}{2}$ ways. Thus

$$\begin{aligned} \mathbb{P}(E_1) &= \frac{\binom{i}{2} \cdot \binom{i-1}{2} \cdots \binom{2}{2} \cdot \binom{n-i}{2} \cdot \binom{n-i-1}{2} \cdots \binom{j+1}{2} \cdot j}{\binom{n}{2} \cdot \binom{n-1}{2} \cdots \binom{j+1}{2}} \\ &= \frac{\binom{i}{2} \cdots \binom{2}{2} \cdot j}{\binom{n}{2} \cdot \binom{n-1}{2} \cdots \binom{n-i+1}{2}} \\ &= \frac{2j}{\binom{n}{i}} \cdot \frac{(i-1)!}{(n-1) \cdots (n-i)}. \end{aligned} \quad (41)$$

Note that this probability does not depend on the choice of the levels k_1, \dots, k_{i-1} . There are $\binom{n-j-1}{i-1}$ possible ways to choose these levels and therefore

$$\mathbb{P}(J = j) = \binom{n-j-1}{i-1} \cdot \mathbb{P}(E_1) = \frac{2j}{\binom{n}{i}} \cdot \frac{(n-j-1)!}{(n-1) \cdots (n-i)(n-j-i)!} \leq \frac{2j}{(n-1)} \cdot \frac{1}{\binom{n}{i}}. \quad (42)$$

Let us now look at the joint distribution of J and J' and consider first the case where $A \cap A' = \emptyset$ and $i + i' < n$. Let us assume first that $j' < j$. If $j > n - i$ or $j' > n - i - i'$ then $\mathbb{P}(J = j, J' = j') = 0$. Otherwise, the $i' - 1$ levels at which mergers in $\mathcal{T}_{A'}$ take place are all different from the $i - 1$ levels at which the mergers in \mathcal{T}_A occur and they are also different from j . At these $i - 1$ (and respectively $i' - 1$) levels pairs of branches from the subtree \mathcal{T}_A (respectively from $\mathcal{T}_{A'}$) are chosen to merge, whereas at the other levels (from level $n - 1$ to level $j' + 1$, excepting j) branches with leaves in $\{1, \dots, n\} \setminus (A \cup A')$ merge. At level j the branch with leaves in A ends through a merger with one of the $j - q$ existing branches with leaves in $\{1, \dots, n\} \setminus (A \cup A')$. At level j' the branch with leaves in A' ends by merging with one of the j' branches with leaves in $\{1, \dots, n\} \setminus (A \cup A')$. Similar as above denote by E_2 the event that the subtrees \mathcal{T}_A and $\mathcal{T}_{A'}$ have their mergers at the fixed levels $n > k_1 > \cdots > k_{i-1} > j$ and $n > k'_1 > \cdots > k'_{i'-1} > j$ respectively. We obtain

$$\begin{aligned} \mathbb{P}(E_2) &= \frac{\binom{i}{2} \cdots \binom{2}{2} \cdot \binom{i'}{2} \cdots \binom{2}{2} \cdot \binom{n-(i+i')}{2} \cdots \binom{j'+1}{2} \cdot (j-q) \cdot j'}{\binom{n}{2} \cdot \binom{n-1}{2} \cdots \binom{j'+1}{2}} \\ &\leq \frac{\binom{i}{2} \cdots \binom{2}{2} \cdot \binom{i'}{2} \cdots \binom{2}{2} \cdot j j'}{\binom{n}{2} \cdot \binom{n-1}{2} \cdots \binom{n-(i+i')+1}{2}} \\ &\leq \frac{4j j'}{\binom{n}{i, i', n-(i+i')}} \cdot \frac{(i-1)!(i'-1)!}{(n-1) \cdots (n-(i+i'))}. \end{aligned} \quad (43)$$

There are $\binom{n-j-1}{i-1} \binom{n-j'-i-1}{i'-1}$ possible ways to first choose the levels k_1, \dots, k_{i-1} and then $k'_1, \dots, k'_{i'-1}$ such that $\{k'_1, \dots, k'_{i'-1}\} \cap \{k_1, \dots, k_{i-1}, j\} = \emptyset$. Thus, since $1 \leq j' < j$,

$$\begin{aligned} \mathbb{P}(J = j, J' = j') &= \binom{n-j-1}{i-1} \binom{n-j'-i-1}{i'-1} \cdot \mathbb{P}(E_2) \\ &\leq \frac{4j j'}{\binom{n}{i, i', n-(i+i')}} \\ &\quad \cdot \frac{(n-j-1)!}{(n-1) \cdots (n-i)(n-j-i)!} \cdot \frac{(n-j'-i-1)!}{(n-i-1) \cdots (n-i-i')(n-j'-i-i')!} \\ &\leq \frac{4j j'}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i, i', n-(i+i')}}. \end{aligned} \quad (44)$$

Observe also that by putting together (44), (43), (41) and (42) we get

$$\mathbb{P}(J = j, J' = j') \leq \binom{n-j-1}{i-1} \frac{\binom{i}{2} \cdots \binom{2}{2}}{\binom{n}{2} \cdots \binom{n-i+1}{2}} \cdot j \cdot \binom{n-j'-1}{i'-1} \frac{\binom{i'}{2} \cdots \binom{2}{2}}{\binom{n-i}{2} \cdots \binom{n-(i+i')+1}{2}} \cdot j'$$

$$\begin{aligned}
&= \mathbb{P}(J = j) \mathbb{P}(J' = j') \cdot \frac{\binom{n}{2} \cdots \binom{n-i'+1}{2}}{\binom{n-i}{2} \cdots \binom{n-i-i'+1}{2}} \\
&\leq \mathbb{P}(J = j) \mathbb{P}(J' = j') \left(1 + \frac{c}{n}\right)
\end{aligned}$$

for a finite constant c depending on i and i' .

In the case $j = j'$, $i + i' < n$ we have that $j = j' \geq 2$. If $j > n - i - i' + 1$ then $\mathbb{P}(J = j, J' = j) = 0$. Otherwise the two branches we consider end at level j by coalescing with one another and therefore we obtain with a similar argument as above that

$$\begin{aligned}
\mathbb{P}(J = j, J' = j) &= \binom{n-j-1}{i-1} \binom{n-j-i}{i'-1} \frac{\binom{i}{2} \cdots \binom{2}{2} \cdot \binom{i'}{2} \cdots \binom{2}{2} \cdot \binom{n-(i+i')}{2} \cdots \binom{j}{2} \cdot 1}{\binom{n}{2} \cdot \binom{n-1}{2} \cdots \binom{j+1}{2}} \\
&= \binom{n-j-1}{i-1} \binom{n-j-i}{i'-1} \frac{\binom{i}{2} \cdots \binom{2}{2} \cdot \binom{i'}{2} \cdots \binom{2}{2}}{\binom{n}{2} \cdots \binom{n-(i+i')+1}{2}} \cdot \binom{j}{2} \\
&= \frac{4 \binom{j}{2}}{\binom{n}{i, i', n-(i+i')}} \cdot \frac{1}{(n-1)(n-2)} \cdot \frac{(n-j-1) \cdots (n-j-i-i'+2)}{(n-3) \cdots (n-i-i')}.
\end{aligned} \tag{45}$$

Using $j \geq 2, i + i' < n$ we get that

$$\mathbb{P}(J = j, J' = j) \leq \frac{2j^2}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i, i', n-(i+i')}}.$$

Also in this case by putting together (45), (41) and (42) we get

$$\begin{aligned}
\mathbb{P}(J = j, J' = j) &\leq \binom{n-j-1}{i-1} \frac{\binom{i}{2} \cdots \binom{2}{2}}{\binom{n}{2} \cdots \binom{n-i+1}{2}} \cdot j \cdot \binom{n-j-1}{i'-1} \frac{\binom{i'}{2} \cdots \binom{2}{2}}{\binom{n-i}{2} \cdots \binom{n-(i+i')+1}{2}} \cdot \frac{j-1}{2} \\
&\leq \mathbb{P}(J = j) \mathbb{P}(J' = j) \cdot \frac{\binom{n}{2} \cdots \binom{n-i'+1}{2}}{\binom{n-i}{2} \cdots \binom{n-i-i'+1}{2}} \\
&\leq \mathbb{P}(J = j) \mathbb{P}(J' = j) \left(1 + \frac{c}{n}\right)
\end{aligned}$$

for a finite constant c depending on i and i' .

Let now $A' \subset A$, $A \neq A'$. For $j' > n - i'$ or $j > n - i$ it holds that $\mathbb{P}(J = j, J' = j) = 0$. Otherwise $\mathcal{T}_{A'}$ is a subtree of \mathcal{T}_A and $j' > j$. If \mathcal{T}_A has its mergers at the levels k_1, \dots, k_{i-1} , then $j' \in \{k_1, \dots, k_{i-1}\}$ and at level j' the remaining branch of $\mathcal{T}_{A'}$ merges with one of the q existing branches with leaves in $A \setminus A'$, where $1 \leq q \leq i - i'$. By similar arguments as above we deduce that

$$\mathbb{P}(J = j, J' = j') \leq \binom{n-j'-1}{i'-1} \binom{n-j-i'-1}{i-i'-1} \frac{\binom{i-i'}{2} \cdots \binom{2}{2} \cdot \binom{i'}{2} \cdots \binom{2}{2}}{\binom{n}{2} \cdots \binom{n-i+1}{2}} \cdot j \cdot (i - i').$$

Replacing j by 1 and j' by 2 in the binomials we obtain

$$\mathbb{P}(J = j, J' = j') \leq 4j(i - i') \cdot \frac{1}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i-i', i', n-i}}.$$

(ii) Let now $i \geq 2$. If $k > n - i$ then $\mathbb{P}(K = k, J = j) = \mathbb{P}(K = k) = 0$. Otherwise the probability that the subtree \mathcal{T}_A has its last merger at level k and its root at level j can be quickly obtained by summing up the equality (41) over all possible ways of choosing the merging levels $n > k_1 > \dots > k_{i-2} > k$. Thus

$$\mathbb{P}(K = k, J = j) = \binom{n-k-1}{i-2} \frac{\binom{i}{2} \cdots \binom{2}{2}}{\binom{n}{2} \cdots \binom{n-i+1}{2}} \cdot j \leq \frac{2ij}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i}}. \tag{46}$$

Summing now over all $1 \leq j < k$ we get that

$$\mathbb{P}(K = k) \leq \sum_{j=1}^k \frac{2ij}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i}} \leq \frac{ik^2}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i}}.$$

As to the last claim of the lemma, $\mathbb{P}(K = k, J = j, K' = k', J' = j') = 0$ if $k = k'$. Without loss of generality assume $k' < k$. Note that the condition $i + i' < n$ implies that $J > 1$ or $J' > 1$. Considering the positions of j, j' and k' we have to distinguish between the cases $j < j', j = j', j' < j < k'$ and $k' < j$. In the first case, one needs to choose the levels k_1, \dots, k_{i-2} and $k'_1, \dots, k'_{i'-2}$ such that

$$\{k'_1, \dots, k'_{i'-2}\} \cap \{k_1, \dots, k_{i-2}, k\} = \emptyset.$$

Since $j < j'$ the branch supporting $\mathcal{T}_{A'}$ (which ends at level j') cannot merge with the branch supporting \mathcal{T}_A and must therefore merge with one of the $j' - 1$ branches left. Thus

$$\mathbb{P}(K = k, J = j, K' = k', J' = j') = \binom{n-k-1}{i-2} \binom{n-k'-i}{i'-2} \frac{\binom{i}{2} \cdots \binom{2}{2} \cdot \binom{i'}{2} \cdots \binom{2}{2}}{\binom{n}{2} \cdots \binom{n-(i+i')+1}{2}} \cdot j \cdot (j' - 1) \quad (47)$$

which by (46) leads to

$$\begin{aligned} \mathbb{P}(K = k, J = j, K' = k', J' = j') &\leq \mathbb{P}(K = k, J = j) \mathbb{P}(K' = k', J' = j') \cdot \frac{\binom{n}{2} \cdots \binom{n-i'+1}{2}}{\binom{n}{2} \cdots \binom{n-(i+i')+1}{2}} \\ &\leq \mathbb{P}(K = k, J = j) \mathbb{P}(K' = k', J' = j') \left(1 + \frac{c}{n}\right) \end{aligned}$$

for a finite constant c depending on i and i' .

In the case $j' < j < k'$ the same argument holds, with the observation that the product $j \cdot (j' - 1)$ on the right-hand side of (47) has to be replaced by $(j - 1) \cdot j'$, whereas in the case $j = j'$ it is $\binom{j}{2}$ that appears instead of the product $j \cdot (j' - 1)$. In the last case, $k' < j$, observe that the intermediate levels k_1, \dots, k_{i-2} and $k'_1, \dots, k'_{i'-2}$ have to be chosen such that

$$\{k'_1, \dots, k'_{i'-2}\} \cap \{k_1, \dots, k_{i-2}, k, j\} = \emptyset$$

and thus

$$\mathbb{P}(K = k, J = j, K' = k', J' = j') = \binom{n-k-1}{i-2} \binom{n-k'-i-1}{i'-2} \frac{\binom{i}{2} \cdots \binom{2}{2} \cdot \binom{i'}{2} \cdots \binom{2}{2}}{\binom{n}{2} \cdots \binom{n-(i+i')+1}{2}} \cdot (j - q) \cdot j'$$

where q denotes as above the number of branches supporting leaves with labels in A' extant at level j (and with which the branch supporting the leaves with labels in A can not merge). By (46) we obtain the desired upper bound. This finishes the proof. \square

The following lemma employs Lemma 4 extensively and will be of use in the proof of Theorem 1.

Lemma 5. *For fixed $i \geq 1$ as $n \rightarrow \infty$*

$$\frac{n}{\log n} \sum_{\substack{A \in \mathcal{P}_n \\ |A|=i}} \left(\frac{1}{J_A} - \mathbb{E}\left(\frac{1}{J_A}\right) \right)^2 \xrightarrow{\mathbb{P}} 1 \quad (48)$$

with $\frac{1}{J_A} = 0$ if $J_A = \infty$. Also

$$\max_{\substack{A \in \mathcal{P}_n \\ |A|=i}} \sqrt{\frac{n}{\log n}} \left| \frac{1}{J_A} - \mathbb{E}\left(\frac{1}{J_A}\right) \right| \xrightarrow{\mathbb{P}} 0. \quad (49)$$

Proof. Remember that by definition $J_A = \infty$ if there exists no branch in the coalescent supporting the leaves with labels in the set A and otherwise $J_A < n$. It thus holds that

$$\frac{n}{\log n} \sum_{\substack{A \in \mathcal{P}_n \\ |A|=i}} \frac{1}{J_A^2} = \frac{n}{\log n} \sum_{\substack{A \in \mathcal{P}_n \\ |A|=i}} \frac{1}{J_A^2} \cdot \mathbf{1}_{\{J_A < n\}}.$$

In what follows we show that

$$\frac{n}{\log n} \sum_{|A|=i} \frac{1}{J_A^2} \cdot \mathbf{1}_{\{J_A < n\}} \xrightarrow{\mathbb{P}} 1. \quad (50)$$

Let us denote by $\mathcal{A} = \mathcal{A}(n, i, a)$ the event that at least one branch of order i ends between level 2 and level $a\sqrt{n}$ in the coalescent, where a is a positive constant:

$$\mathcal{A} = \mathcal{A}(n, i, a) := \{J_A < a\sqrt{n} \quad \text{for some } A \in \mathcal{P}_n, |A| = i\}.$$

Note by Lemma 4 that

$$\mathbb{P}(\mathcal{A}) \leq \sum_{|A|=i} \mathbb{P}(J_A < a\sqrt{n}) \leq \sum_{|A|=i} \sum_{1 \leq j < a\sqrt{n}} \frac{2j}{(n-1)} \cdot \frac{1}{\binom{n}{i}} \leq 2a^2. \quad (51)$$

Since $\sum_{|A|=i} \frac{1}{J_A^2} \cdot \mathbf{1}_{\{J_A < n\}} \neq \sum_{|A|=i} \frac{1}{J_A^2} \cdot \mathbf{1}_{\{a\sqrt{n} \leq J_A < n\}}$ only on the event \mathcal{A} , which for small a has arbitrary small probability, in order to prove (50) it suffices to show that for all $a > 0$

$$\frac{n}{\log n} \sum_{|A|=i} \frac{1}{J_A^2} \cdot \mathbf{1}_{\{a\sqrt{n} \leq J_A < n\}} \xrightarrow{\mathbb{P}} 1 \quad \text{as } n \rightarrow \infty.$$

Using (42) we obtain that for all $a > 0$

$$\begin{aligned} \mathbb{E}\left(\frac{n}{\log n} \sum_{|A|=i} \frac{1}{J_A^2} \cdot \mathbf{1}_{\{a\sqrt{n} \leq J_A < n\}}\right) &= \frac{n}{\log n} \sum_{|A|=i} \sum_{a\sqrt{n} \leq j < n} \frac{1}{j^2} \cdot \frac{2j}{\binom{n}{i}} \cdot \frac{(n-j-1) \cdots (n-j-i+1)}{(n-1) \cdots (n-i)} \\ &= \frac{n}{(n-1) \log n} \sum_{a\sqrt{n} \leq j < n} \frac{2}{j} \cdot \frac{(n-2-(j-1)) \cdots (n-i-(j-1))}{(n-2) \cdots (n-i)} \\ &\rightarrow 1 \end{aligned} \quad (52)$$

as $n \rightarrow \infty$.

For the variance we obtain

$$\begin{aligned} \mathbb{V}\left(\frac{n}{\log n} \sum_{|A|=i} \frac{1}{J_A^2} \cdot \mathbf{1}_{\{a\sqrt{n} \leq J_A < n\}}\right) &\leq \frac{n^2}{\log^2 n} \sum_A \mathbb{E}\left(\frac{1}{J_A^4} \cdot \mathbf{1}_{\{a\sqrt{n} \leq J_A < n\}}\right) \\ &\quad + \frac{n^2}{\log^2 n} \sum_{A \cap A' = \emptyset} \text{COV}\left(\frac{1}{J_A^2} \cdot \mathbf{1}_{\{a\sqrt{n} \leq J_A < n\}}, \frac{1}{J_{A'}^2} \cdot \mathbf{1}_{\{a\sqrt{n} \leq J_{A'} < n\}}\right). \end{aligned} \quad (53)$$

Note that for $A \cap A' \notin \{A, A', \emptyset\}$, $|A| = |A'| = i$ either J_A or $J_{A'} = \infty$ and thus the corresponding covariances are less than or equal to 0.

By Lemma 4 it holds for $A \cap A' = \emptyset$ that

$$\begin{aligned} \text{COV}\left(\frac{1}{J_A^2} \cdot \mathbf{1}_{\{a\sqrt{n} \leq J_A < n\}}, \frac{1}{J_{A'}^2} \cdot \mathbf{1}_{\{a\sqrt{n} \leq J_{A'} < n\}}\right) \\ = \sum_{a\sqrt{n} \leq j, j' < n} \frac{1}{j^2 j'^2} \cdot \left(\mathbb{P}(J_A = j, J_{A'} = j') - \mathbb{P}(J_A = j) \mathbb{P}(J_{A'} = j')\right) \end{aligned}$$

$$\begin{aligned}
&\ll \sum_{a\sqrt{n} \leq j, j' < n} \frac{1}{j^2 j'^2} \cdot \mathbb{P}(J_A = j) \mathbb{P}(J_{A'} = j') \cdot \frac{1}{n}. \\
&\ll \sum_{a\sqrt{n} \leq j, j' < n} \frac{1}{j j'} \cdot \frac{1}{n^2} \cdot \frac{1}{\binom{n}{i}^2} \cdot \frac{1}{n} \\
&\ll \frac{(\log n)^2}{n^3} \cdot \frac{1}{\binom{n}{i}^2},
\end{aligned}$$

for a constant c depending on i and i' . Plugging this in (53) and using again Lemma 4 we obtain that

$$\begin{aligned}
&\mathbb{V}\left(\frac{n}{\log n} \sum_{|A|=i} \frac{1}{J_A^2} \cdot \mathbf{1}_{\{a\sqrt{n} \leq J_A < n\}}\right) \\
&\ll \frac{n^2}{(\log n)^2} \sum_A \sum_{a\sqrt{n} \leq j < n} \frac{1}{j^4} \cdot \frac{2j}{n-1} \cdot \frac{1}{\binom{n}{i}} + \frac{n^2}{(\log n)^2} \sum_{A \cap A' = \emptyset} \frac{(\log n)^2}{n^3} \cdot \frac{1}{\binom{n}{i}^2} \\
&\ll \frac{n^2}{(\log n)^2} \left(\frac{1}{a^2 n^2} + \frac{(\log n)^2}{n^3} \right) \\
&\ll \frac{1}{(\log n)^2} + \frac{1}{n}
\end{aligned}$$

and thus

$$\mathbb{V}\left(\frac{n}{\log n} \sum_{|A|=i} \frac{1}{J_A^2} \cdot \mathbf{1}_{\{a\sqrt{n} \leq J_A < n\}}\right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

This together with (52) and (51) proves (50).

From Lemma 4

$$\mathbb{E}\left(\frac{1}{J_A}\right) \leq \sum_{j=1}^{n-1} \frac{1}{j} \cdot \frac{2j}{(n-1)} \cdot \frac{1}{\binom{n}{i}} = \frac{2}{\binom{n}{i}} \quad (54)$$

and thus for $i \geq 1$

$$\frac{n}{\log n} \sum_{|A|=i} \mathbb{E}\left(\frac{1}{J_A}\right)^2 \ll \frac{n}{\log n} \sum_A \frac{1}{\binom{n}{i}^2} = \frac{n}{\log n} \cdot \frac{1}{\binom{n}{i}} \rightarrow 0$$

as $n \rightarrow \infty$. Because of (50) it follows by the Cauchy-Schwarz inequality that

$$\frac{n}{\log n} \sum_{|A|=i} 2\left(\frac{1}{J_A}\right) \cdot \mathbb{E}\left(\frac{1}{J_A}\right) \leq \frac{2n}{\log n} \sqrt{\sum_A \frac{1}{J_A^2}} \cdot \sqrt{\sum_A \mathbb{E}\left(\frac{1}{J_A}\right)^2} \xrightarrow{\mathbb{P}} 0$$

as $n \rightarrow \infty$ and thus (48) holds.

Finally observe that by (54)

$$\left| \frac{1}{J_A} - \mathbb{E}\left(\frac{1}{J_A}\right) \right| \leq \frac{1}{J_A} + \frac{2}{n} \leq \frac{1}{a\sqrt{n}} + \frac{2}{n}$$

on the event \mathcal{A}^c . Thus the claim (49) is a direct consequence of (51). \square

5 Contribution of big families

In this section we show that the branches of big orders from the coalescent at time 0 make only a negligible contribution to the external length at a later time h when the number of individuals in the population gets large.

Recall the definition of $R_h^{n,r}$ given in (4) and let

$$a_{n,h} := \left(\log \left(\frac{h+2}{h} \right) \right)^{-1} \log n. \quad (55)$$

Then we can write

$$R_h^{n,r} = \tilde{R}_h^{n,r} + \bar{R}_h^{n,r},$$

where

$$\tilde{R}_h^{n,r} := \sum_{\substack{A \in \mathcal{P}_n \\ r < |A| \leq a_{n,h}}} \mathbf{1}_{\mathcal{F}_A} \cdot L_A, \quad \bar{R}_h^{n,r} := \sum_{\substack{A \in \mathcal{P}_n \\ a_{n,h} < |A| < n}} \mathbf{1}_{\mathcal{F}_A} \cdot L_A.$$

Proposition 2. *For each $h \geq 0$ there exists a sequence $\{\varepsilon(r)\}_{r \in \mathbb{N}}$ with $\varepsilon(r) = o(1)$ when $r \rightarrow \infty$, such that*

$$\mathbb{V}(\tilde{R}_h^{n,r}) \leq \varepsilon(r) \frac{\log n}{n}$$

holds. Moreover $\mathbb{E}(\bar{R}_h^{n,r}) = O(\frac{1}{n})$.

Proof. First we prove the second claim. Since the collection of random variables $\{\mathbf{1}_{\mathcal{F}_A}\}_{A \in \mathcal{P}_n}$ is independent of the collection of random variables $\{L_A\}_{A \in \mathcal{P}_n}$ due to the Poissonian structure embedded in the Moran model, it holds that

$$\mathbb{E} \left(\sum_{|A| > a_{n,h}} \mathbf{1}_{\mathcal{F}_A} \cdot L_A \right) = \sum_{i > a_{n,h}} \sum_{|A|=i} \mathbb{E}(\mathbf{1}_{\mathcal{F}_A}) \cdot \mathbb{E}(L_A).$$

Note that with the notation introduced in Section 3 it holds that

$$\mathcal{F}_A = \{Z_h^A = 1\}$$

and thus by (18) for $|A| = i$

$$p_{i,h}^n := \mathbb{P}(\mathcal{F}_A) \leq 2i \cdot \left(\frac{h}{h+2} \right)^{i-1}.$$

By (2) it holds for $|A| = i$

$$\mathbb{E}(L_A) = \frac{1}{\binom{n}{i}} \cdot \mathbb{E}(\mathcal{L}^{n,i}) = \frac{1}{\binom{n}{i}} \cdot \frac{2}{i}. \quad (56)$$

We obtain that

$$\begin{aligned} \mathbb{E}(\bar{R}_h^{n,r}) &= \mathbb{E} \left(\sum_{|A| > a_{n,h}} \mathbf{1}_{\mathcal{F}_A} \cdot L_A \right) \leq \sum_{i > a_{n,h}} \binom{n}{i} \cdot 2i \cdot \left(\frac{h}{h+2} \right)^{i-1} \cdot \frac{2}{i} \cdot \frac{1}{\binom{n}{i}} \\ &\ll \left(\frac{h}{h+2} \right)^{(\log(\frac{h+2}{h}))^{-1} \log n} \\ &= \frac{1}{n}. \end{aligned}$$

Thus $\mathbb{E}(\bar{R}_h^{n,r}) = O(\frac{1}{n})$.

We now proceed to proving the first claim of the Proposition. For the variance of $\tilde{R}_h^{n,r}$ it holds that

$$\begin{aligned} \mathbb{V}(\tilde{R}_h^{n,r}) &= \mathbb{V} \left(\sum_{r < |A| \leq a_{n,h}} \mathbf{1}_{\mathcal{F}_A} \cdot L_A \right) \\ &= \mathbb{V} \left(\sum_{r < |A| \leq a_{n,h}} (\mathbf{1}_{\mathcal{F}_A} - \mathbb{E}(\mathbf{1}_{\mathcal{F}_A})) \cdot L_A + \sum_{r < |A| \leq a_{n,h}} \mathbb{E}(\mathbf{1}_{\mathcal{F}_A}) \cdot L_A \right) \end{aligned}$$

$$\begin{aligned}
&= \mathbb{V}\left(\sum_{r < |A| \leq a_{n,h}} (\mathbf{1}_{\mathcal{F}_A} - \mathbb{E}(\mathbf{1}_{\mathcal{F}_A})) \cdot L_A\right) + \mathbb{V}\left(\sum_{r < |A| \leq a_{n,h}} \mathbb{E}(\mathbf{1}_{\mathcal{F}_A}) \cdot L_A\right) \\
&\quad + \mathbb{COV}\left(\sum_{r < |A| \leq a_{n,h}} (\mathbf{1}_{\mathcal{F}_A} - \mathbb{E}(\mathbf{1}_{\mathcal{F}_A})) \cdot L_A, \sum_{r < |A| \leq a_{n,h}} \mathbb{E}(\mathbf{1}_{\mathcal{F}_A}) \cdot L_A\right).
\end{aligned}$$

Taking into account the independence of the collections of random variables $\{\mathbf{1}_{\mathcal{F}_A}\}_{A \in \mathcal{P}_n}$ and $\{L_A\}_{A \in \mathcal{P}_n}$, the last term on the right-hand side above is equal to zero. Therefore, using the fact that $\mathbb{E}(\mathbf{1}_{\mathcal{F}_A})$ depends on the set A only through its cardinality i and again from independence, we have that

$$\begin{aligned}
\mathbb{V}(\tilde{R}_h^{n,r}) &= \mathbb{E}\left(\left(\sum_{r < |A| \leq a_{n,h}} (\mathbf{1}_{\mathcal{F}_A} - \mathbb{E}(\mathbf{1}_{\mathcal{F}_A})) \cdot L_A\right)^2\right) + \mathbb{V}\left(\sum_{r < |A| \leq a_{n,h}} \mathbb{E}(\mathbf{1}_{\mathcal{F}_A}) \cdot L_A\right) \\
&= \sum_{r < |A|, |A'| \leq a_{n,h}} \mathbb{E}\left((\mathbf{1}_{\mathcal{F}_A} - \mathbb{E}(\mathbf{1}_{\mathcal{F}_A}))(\mathbf{1}_{\mathcal{F}_{A'}} - \mathbb{E}(\mathbf{1}_{\mathcal{F}_{A'}}))\right) \cdot \mathbb{E}(L_A L_{A'}) + \mathbb{V}\left(\sum_{r < i \leq a_{n,h}} p_{i,h}^n \cdot \mathcal{L}^{n,i}\right).
\end{aligned} \tag{57}$$

We start by evaluating the first term on the right-hand side. To this aim note that if A and A' are such that $A \cap A' \notin \{\emptyset, A, A'\}$, then there cannot be two branches in the coalescent tree, one supporting the leaves in A and the other one supporting the leaves in A' and therefore in this case the product $L_A L_{A'}$ is equal to zero. Thus, suppressing in the notation the restrictions on the cardinalities of the sets, it holds that

$$\begin{aligned}
&\sum_{A, A' \in \mathcal{P}_n} \mathbb{E}\left((\mathbf{1}_{\mathcal{F}_A} - \mathbb{E}(\mathbf{1}_{\mathcal{F}_A}))(\mathbf{1}_{\mathcal{F}_{A'}} - \mathbb{E}(\mathbf{1}_{\mathcal{F}_{A'}}))\right) \cdot \mathbb{E}(L_A L_{A'}) \\
&\leq 2 \sum_{A' \subsetneq A} \mathbb{E}\left(\mathbf{1}_{\mathcal{F}_A} \mathbf{1}_{\mathcal{F}_{A'}}\right) \cdot \mathbb{E}(L_A L_{A'}) \\
&\quad + \sum_{A \cap A' = \emptyset} \mathbb{COV}(\mathbf{1}_{\mathcal{F}_A}, \mathbf{1}_{\mathcal{F}_{A'}}) \cdot \mathbb{E}(L_A L_{A'}) + \sum_A \mathbb{V}(\mathbf{1}_{\mathcal{F}_A}) \cdot \mathbb{E}(L_A^2).
\end{aligned} \tag{58}$$

Using the definition (40) of L_A we obtain that

$$\begin{aligned}
\mathbb{E}(L_A L_{A'}) &= \mathbb{E}\left(\mathbb{E}(L_A L_{A'} \mid J_A, J_{A'}, K_A, K_{A'})\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\sum_{j=J_A+1}^{K_A} X_j \cdot \sum_{l=J_{A'}+1}^{K_{A'}} X_l \mid J_A, J_{A'}, K_A, K_{A'}\right)\right) \\
&= \mathbb{E}\left(\sum_{j=J_A+1}^{K_A} \sum_{l=J_{A'}+1}^{K_{A'}} \mathbb{E}(X_j X_l)\right).
\end{aligned}$$

The exponential inter-coalescence times X_j and X_l are independent if $j \neq l$ and in this case $\mathbb{E}(X_j X_l) = \frac{2}{j(j-1)} \frac{2}{l(l-1)}$, whereas if $j = l$ it holds that $\mathbb{E}(X_j X_l) = \mathbb{E}(X_j^2) = 2 \left(\frac{2}{j(j-1)}\right)^2 = 2 \frac{2}{j(j-1)} \frac{2}{l(l-1)}$. We thus obtain that

$$\begin{aligned}
\mathbb{E}(L_A L_{A'}) &\leq 2 \mathbb{E}\left(\sum_{j=J_A+1}^{K_A} \sum_{l=J_{A'}+1}^{K_{A'}} \frac{2}{j(j-1)} \frac{2}{l(l-1)}\right) \\
&= 2 \mathbb{E}\left(\sum_{j=J_A+1}^{K_A} \frac{2}{j(j-1)} \cdot \sum_{l=J_{A'}+1}^{K_{A'}} \frac{2}{l(l-1)}\right) \\
&= 2 \mathbb{E}\left(\left(\frac{2}{J_A} - \frac{2}{K_A}\right)\left(\frac{2}{J_{A'}} - \frac{2}{K_{A'}}\right)\right) \\
&\leq 8 \mathbb{E}\left(\frac{1}{J_A J_{A'}}\right).
\end{aligned} \tag{59}$$

In order to bound the first term on the right-hand side of (58) observe first that by (18)

$$\mathbb{E}(\mathbf{1}_{\mathcal{F}_A} \mathbf{1}_{\mathcal{F}_{A'}}) \leq \mathbb{E}(\mathbf{1}_{\mathcal{F}_A}) \leq 2i \left(\frac{h}{h+2} \right)^{i-1}.$$

Now, using (59) together with the fact that $A' \subsetneq A$ implies that $J_{A'} > J_A$ and Lemma 4 we obtain

$$\begin{aligned} & \sum_{A' \subsetneq A} \mathbb{E}(\mathbf{1}_{\mathcal{F}_A} \mathbf{1}_{\mathcal{F}_{A'}}) \cdot \mathbb{E}(L_A L_{A'}) \\ & \leq 16 \sum_{i' < i} \binom{n}{i-i', i', n-i} \cdot i \left(\frac{h}{h+2} \right)^{i-1} \cdot \sum_{1 \leq j < j' < n} \frac{1}{jj'} \mathbb{P}(J_A = j, J_{A'} = j') \\ & \leq 16 \sum_{i' < i} \binom{n}{i-i', i', n-i} \cdot i \left(\frac{h}{h+2} \right)^{i-1} \cdot \sum_{1 \leq j < j' < n} \frac{1}{jj'} \frac{4j(i-i')}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i-i', i', n-i}} \\ & \leq 64 \sum_{i' < i} i \left(\frac{h}{h+2} \right)^{i-1} \cdot \frac{i-i'}{(n-1)(n-2)} \sum_{1 \leq j < j' < n} \frac{1}{j'} \\ & \ll \frac{1}{n} \cdot \sum_{i' < i} i(i-i') \left(\frac{h}{h+2} \right)^{i-1} \\ & \ll \frac{1}{n} \cdot \sum_{i > r} i^3 \left(\frac{h}{h+2} \right)^{i-1} \\ & \ll \frac{1}{n}. \end{aligned} \tag{60}$$

For bounding the second term on the right-hand side of (58) observe that for $A \cap A' = \emptyset$ by (59) and Lemma 4 it holds that

$$\begin{aligned} \mathbb{E}(L_A L_{A'}) & \leq 8 \sum_{j, j' \geq 1} \frac{1}{jj'} \mathbb{P}(J_A = j, J_{A'} = j') \\ & \leq 32 \sum_{j, j' \geq 1} \frac{1}{jj'} \frac{jj'}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i, i', n-(i+i')}} \\ & \ll \frac{1}{\binom{n}{i, i', n-(i+i')}} \end{aligned}$$

and therefore by (39) we obtain that for some $\gamma < 1$ depending on h it holds

$$\sum_{A \cap A' = \emptyset} \mathbb{COV}(\mathbf{1}_{\mathcal{F}_A}, \mathbf{1}_{\mathcal{F}_{A'}}) \cdot \mathbb{E}(L_A L_{A'}) \ll \sum_{A \cap A' = \emptyset} \frac{1}{n} \cdot \gamma^{i+i'} \cdot \frac{1}{\binom{n}{i, i', n-(i+i')}} \ll \frac{1}{n} \sum_{i, i' > r} \gamma^{i+i'} \ll \frac{1}{n}. \tag{61}$$

We now turn to the last term on the right-hand side of (58), namely $\sum_A \mathbb{V}(\mathbf{1}_{\mathcal{F}_A}) \cdot \mathbb{E}(L_A^2)$. From (18) it holds that

$$\mathbb{V}(\mathbf{1}_{\mathcal{F}_A}) \leq \mathbb{E}(\mathbf{1}_{\mathcal{F}_A}) \leq 2i \left(\frac{h}{h+2} \right)^{i-1}$$

and by (59) and Lemma 4

$$\mathbb{E}(L_A^2) \leq 8 \sum_{1 \leq j < n} \frac{1}{j^2} \mathbb{P}(J_A = j) \leq 16 \sum_{1 \leq j < n} \frac{1}{j^2} \frac{j}{(n-1)} \cdot \frac{1}{\binom{n}{i}} \ll \frac{1}{\binom{n}{i}} \cdot \frac{\log n}{n}.$$

This leads to

$$\sum_A \mathbb{V}(\mathbf{1}_{\mathcal{F}_A}) \cdot \mathbb{E}(L_A^2) \ll \sum_{r < i \leq a_{n,h}} \binom{n}{i} \cdot i \left(\frac{h}{h+2} \right)^{i-1} \cdot \frac{1}{\binom{n}{i}} \cdot \frac{\log n}{n} \leq \varepsilon_1(r) \frac{\log n}{n} \quad (62)$$

where $\varepsilon_1(r) := \sum_{i=r+1}^{\infty} i \left(\frac{h}{h+2} \right)^{i-1}$ has the property that

$$\varepsilon_1(r) = o(1) \quad \text{when } r \rightarrow \infty. \quad (63)$$

Putting together (58) and (60) - (62) we obtain for the first term on the right-hand side of (57) that

$$\sum_{r < |A|, |A'| \leq a_{n,h}} \mathbb{E}((\mathbf{1}_{\mathcal{F}_A} - \mathbb{E}(\mathbf{1}_{\mathcal{F}_A}))(\mathbf{1}_{\mathcal{F}_{A'}} - \mathbb{E}(\mathbf{1}_{\mathcal{F}_{A'}}))) \cdot \mathbb{E}(L_A L_{A'}) \ll \left(\varepsilon_1(r) \frac{\log n}{n} + \frac{1}{n} \right). \quad (64)$$

In order to obtain the claim of the Proposition we are left to bound the second term on the right-hand side of (57). For this term it holds that

$$\mathbb{V}\left(\sum_{r < i \leq a_{n,h}} p_{i,h}^n \cdot \mathcal{L}^{n,i}\right) = \sum_{r < i \leq a_{n,h}} (p_{i,h}^n)^2 \cdot \mathbb{V}(\mathcal{L}^{n,i}) + \sum_{\substack{r < i, i' \leq a_{n,h} \\ i \neq i'}} p_{i,h}^n p_{i',h}^n \cdot \mathbb{COV}(\mathcal{L}^{n,i}, \mathcal{L}^{n,i'}). \quad (65)$$

The variances and covariances of the internal lengths of different orders can be easily obtained from the results of Fu [13] on the variances and covariances of the numbers $M_i(n)$ of mutations carried by exactly i individuals in a population of size n evolving according to the Moran model under the infinitely many sites mutation model. In this setting mutations are modelled as points of a Poisson process with constant rate $\frac{\phi}{2}$ per unit length on the branches of the coalescent tree. Therefore it holds that

$$(M_i(n) \mid \mathcal{L}^{n,i}) \sim \text{Poisson} \left(\frac{\phi}{2} \mathcal{L}^{n,i} \right).$$

We thus obtain by the law of total variance and of total covariance respectively that

$$\begin{aligned} \mathbb{V}(M_i(n)) &= \mathbb{V}(\mathbb{E}(M_i(n) \mid \mathcal{L}^{n,i})) + \mathbb{E}(\mathbb{V}(M_i(n) \mid \mathcal{L}^{n,i})) \\ &= \mathbb{V}\left(\frac{\phi}{2} \mathcal{L}^{n,i}\right) + \mathbb{E}\left(\frac{\phi}{2} \mathcal{L}^{n,i}\right) \\ &= \frac{\phi^2}{4} \mathbb{V}(\mathcal{L}^{n,i}) + \frac{\phi}{i} \end{aligned} \quad (66)$$

and due to the independence ensured by the Poisson structure

$$\begin{aligned} \mathbb{COV}(M_i(n), M_{i'}(n)) &= \mathbb{COV}(\mathbb{E}(M_i(n) \mid \mathcal{L}^{n,i}, \mathcal{L}^{n,i'}), \mathbb{E}(M_{i'}(n) \mid \mathcal{L}^{n,i}, \mathcal{L}^{n,i'})) \\ &\quad + \mathbb{E}(\mathbb{COV}(M_i(n), M_{i'}(n) \mid \mathcal{L}^{n,i}, \mathcal{L}^{n,i'})) \\ &= \mathbb{COV}\left(\frac{\phi}{2} \mathcal{L}^{n,i}, \frac{\phi}{2} \mathcal{L}^{n,i'}\right) + 0 \\ &= \frac{\phi^2}{4} \mathbb{COV}(\mathcal{L}^{n,i}, \mathcal{L}^{n,i'}). \end{aligned} \quad (67)$$

The results of [13] say that

$$\mathbb{V}(M_i(n)) = \phi^2 \sigma_{ii} + \frac{\phi}{i} \quad \text{and} \quad \mathbb{COV}(M_i(n), M_{i'}(n)) = \phi^2 \sigma_{ii'}, \quad (68)$$

where in particular for $i < \frac{n}{2}$

$$\sigma_{ii} = \beta_n(i+1)$$

and for $i > i'$, $i + i' < \frac{n}{2}$

$$\sigma_{ii'} = \frac{\beta_n(i+1) - \beta_n(i)}{2}$$

with $h_n = \sum_{j=1}^{n-1} \frac{1}{j}$ and

$$\beta_n(i) = \frac{2n}{(n-i+1)(n-i)}(h_{n+1} - h_i) - \frac{2}{n-i}.$$

Therefore from (66), (67) and (68) it follows that

$$\mathbb{V}(\mathcal{L}^{n,i}) = 4\sigma_{ii} \quad \text{and} \quad \mathbb{COV}(\mathcal{L}^{n,i}, \mathcal{L}^{n,i'}) = 4\sigma_{ii'}.$$

Turning now to (65) note from the definition of $a_{n,h}$ given in (55) that the indices i and i' run only up to values of the order $\log n$. Therefore, for large n , it holds that $i < \frac{n}{2}$ and $i + i' < \frac{n}{2}$ and hence we need the values of σ_{ii} and $\sigma_{ii'}$ only in these particular cases. For $i < \frac{n}{2}$ the following bound holds

$$\sigma_{ii} = \beta_n(i+1) \leq \frac{2n h_{n+1}}{(n-i+1)(n-i)} \leq c \cdot \frac{\log n}{n},$$

for c a finite constant independent of i . Also $\sigma_{ii'} \leq 0$ for $i > i'$, $i + i' < \frac{n}{2}$ (see (36) in [13]). Hence using (18), (65) becomes

$$\mathbb{V}\left(\sum_{r < i \leq a_{n,h}} p_{i,h}^n \cdot \mathcal{L}^{n,i}\right) \leq \sum_{r < i \leq a_{n,h}} c \cdot 4i^2 \left(\frac{h}{h+2}\right)^{2i-2} \cdot \frac{\log n}{n} \leq \varepsilon_2(r) \frac{\log n}{n}, \quad (69)$$

where $\varepsilon_2(r) := c \sum_{i=r+1}^{\infty} 4i^2 \left(\frac{h}{h+2}\right)^{2i-2}$ has the property that

$$\varepsilon_2(r) = o(1) \quad \text{when } r \rightarrow \infty. \quad (70)$$

Putting now together (57), (63), (64), (69) and (70) we obtain the claim. \square

6 Proof of Theorem 1

6.1 Preliminaries

Proposition 3. *For any $r \in \mathbb{N}$ it holds that*

$$\sum_{1 \leq |A| \leq r} (\mathbf{1}_{\mathcal{F}_A} \cdot L_A - \mathbb{E}(\mathbf{1}_{\mathcal{F}_A} \cdot L_A)) = \sum_{1 \leq |A| \leq r} \mathbf{1}_{\mathcal{F}_A} \left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right) \right) + O_P(n^{-1/2}) \quad (71)$$

and also

$$\sum_{1 \leq |A| \leq r} (L_A - \mathbb{E}(L_A)) = \sum_{1 \leq |A| \leq r} \left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right) \right) + O_P(n^{-1/2}). \quad (72)$$

Proof. We start by making the observation that for any $r \in \mathbb{N}$ the following claim holds:

$$\sum_{1 \leq |A| \leq r} (\mathbf{1}_{\mathcal{F}_A} \cdot L_A - \mathbb{E}(\mathbf{1}_{\mathcal{F}_A} \cdot L_A)) = \sum_{1 \leq |A| \leq r} \mathbf{1}_{\mathcal{F}_A} \cdot (L_A - \mathbb{E}(L_A)) + O_P(n^{-1/2}). \quad (73)$$

Indeed, since the collections of random variables $\{\mathbf{1}_{\mathcal{F}_A}\}_{A \in \mathcal{P}_n}$ and $\{L_A\}_{A \in \mathcal{P}_n}$ are independent we have that

$$\sum_{1 \leq |A| \leq r} (\mathbf{1}_{\mathcal{F}_A} \cdot L_A - \mathbb{E}(\mathbf{1}_{\mathcal{F}_A} \cdot L_A)) = \sum_{1 \leq |A| \leq r} \mathbf{1}_{\mathcal{F}_A} \cdot (L_A - \mathbb{E}(L_A)) + \sum_{1 \leq |A| \leq r} (\mathbf{1}_{\mathcal{F}_A} - \mathbb{E}(\mathbf{1}_{\mathcal{F}_A})) \cdot \mathbb{E}(L_A).$$

Using (39) and (56) we obtain

$$\begin{aligned} \mathbb{V}\left(\sum_{1 \leq |A| \leq r} (\mathbf{1}_{\mathcal{F}_A} - \mathbb{E}(\mathbf{1}_{\mathcal{F}_A})) \cdot \mathbb{E}(L_A)\right) &= \sum_{1 \leq |A|, |A'| \leq r} \text{COV}(\mathbf{1}_{\mathcal{F}_A}, \mathbf{1}_{\mathcal{F}_{A'}}) \cdot \mathbb{E}(L_A) \mathbb{E}(L_{A'}) \\ &\ll \sum_{i, i'=1}^r \sum_{\substack{|A|=i, |A'|=i' \\ A \cap A' = \emptyset}} \frac{1}{n} \cdot \frac{1}{\binom{n}{i} \binom{n}{i'}} + \sum_{i, i'=1}^r \sum_{\substack{|A|=i, |A'|=i' \\ A \cap A' \neq \emptyset}} \frac{1}{\binom{n}{i} \binom{n}{i'}}. \end{aligned}$$

We now use the fact that the number of non-disjoint sets with cardinalities i and i' is

$$\left| \{A, A' \in \mathcal{P}_n, |A| = i, |A'| = i' \text{ such that } A \cap A' \neq \emptyset\} \right| \leq \binom{n}{i} \cdot i \binom{n}{i' - 1} \ll \frac{1}{n} \binom{n}{i} \binom{n}{i'} \quad (74)$$

and obtain that

$$\mathbb{V}\left(\sum_{1 \leq |A| \leq r} (\mathbf{1}_{\mathcal{F}_A} - \mathbb{E}(\mathbf{1}_{\mathcal{F}_A})) \cdot \mathbb{E}(L_A)\right) \ll \frac{1}{n}.$$

Thus (73) is true.

Let us now introduce some more notation. For $i \in \mathbb{N}$ let

$$\mathcal{G}_i := \sigma\left(J_A, K_A : A \in \mathcal{P}_n, |A| = i\right)$$

be the σ -algebra containing the information about the final and initial levels of the branches of order i in the coalescent at time 0 and let

$$\mathcal{G} := \sigma\left(\bigcup_{i=1}^{n-1} \mathcal{G}_i\right). \quad (75)$$

The σ -algebra \mathcal{G} contains the whole information on the structure of the coalescent tree (up to the exponential times). The conditional expectation of the length of the branch supporting the leaves with labels in A is

$$\mathbb{E}(L_A | \mathcal{G}) = \mathbb{E}\left(\sum_{j=J_A+1}^{K_A} X_j | \mathcal{G}\right) = \sum_{j=J_A+1}^{K_A} \mathbb{E}(X_j) = \frac{2}{J_A} - \frac{2}{K_A}. \quad (76)$$

We next show that by replacing the exponential times X_j in the lengths L_A appearing on the right-hand side of (73) by their expectations leads to a negligible error as the total population size n tends to infinity. In other words, the randomness brought in by the inter-coalescent times can be neglected for big population sizes. The information on the tree structure is contained in \mathcal{G} and therefore the observation we just made amounts to saying that for $i \geq 1$

$$\sum_{|A|=i} \mathbf{1}_{\mathcal{F}_A} L_A = \sum_{|A|=i} \mathbf{1}_{\mathcal{F}_A} \mathbb{E}(L_A | \mathcal{G}) + O_P(n^{-1/2}). \quad (77)$$

Indeed since the collection of random variables $\{\mathbf{1}_{\mathcal{F}_A}\}_{A \in \mathcal{P}_n}$ is independent of $\{L_A\}_{A \in \mathcal{P}_n}$ and independent of the σ -algebra \mathcal{G} , it holds that

$$\begin{aligned} &\mathbb{E}\left(\left(\sum_{|A|=i} \mathbf{1}_{\mathcal{F}_A} (L_A - \mathbb{E}(L_A | \mathcal{G}))\right)^2 | \mathcal{G}\right) \\ &= \mathbb{E}\left(\sum_{|A|=|A'|=i} \mathbf{1}_{\mathcal{F}_A} \mathbf{1}_{\mathcal{F}_{A'}} (L_A - \mathbb{E}(L_A | \mathcal{G}))(L_{A'} - \mathbb{E}(L_{A'} | \mathcal{G})) | \mathcal{G}\right) \end{aligned}$$

$$= \sum_{|A|=|A'|=i} \mathbb{E}(\mathbf{1}_{\mathcal{F}_A} \mathbf{1}_{\mathcal{F}_{A'}}) \text{COV}(L_A, L_{A'} \mid \mathcal{G}).$$

Note that if $A \cap A' \notin \{A, A', \emptyset\}$ then there cannot be two branches in the coalescent, one supporting the leaves with labels in A and one the leaves with labels in A' and therefore in such case L_A or $L_{A'}$ is (by definition) equal to 0 and thus $\mathbb{E}(L_A L_{A'} \mid \mathcal{G}) = 0$ and $\text{COV}(L_A, L_{A'} \mid \mathcal{G}) \leq 0$. Using now (40) we obtain that

$$\begin{aligned} & \mathbb{E}\left(\left(\sum_{|A|=i} \mathbf{1}_{\mathcal{F}_A} (L_A - \mathbb{E}(L_A \mid \mathcal{G}))\right)^2 \mid \mathcal{G}\right) \\ & \leq \sum_A \mathbb{V}\left(\sum_{j=J_A+1}^{K_A} X_j \mid \mathcal{G}\right) + \sum_{A \cap A' = \emptyset} \left(\text{COV}\left(\sum_{j=J_A+1}^{K_A} X_j, \sum_{j'=J_{A'}+1}^{K_{A'}} X_{j'} \mid \mathcal{G}\right)\right)^+ \\ & \leq \sum_A \sum_{j=J_A+1}^{K_A} \mathbb{V}(X_j) + \sum_{A \cap A' = \emptyset} \sum_{j=J_A+1}^{K_A} \sum_{j'=J_{A'}+1}^{K_{A'}} \left(\text{COV}(X_j, X_{j'})\right)^+. \end{aligned}$$

Recall that the inter-coalescent times X_j are independent and exponentially distributed with parameter $\binom{j}{2}$. Therefore

$$\begin{aligned} \mathbb{E}\left(\left(\sum_A \mathbf{1}_{\mathcal{F}_A} (L_A - \mathbb{E}(L_A \mid \mathcal{G}))\right)^2 \mid \mathcal{G}\right) & \leq \sum_A \sum_{j=J_A+1}^{K_A} \mathbb{V}(X_j) + \sum_{A \cap A' = \emptyset} \sum_{j=J_A \vee J_{A'}+1}^{K_A \wedge K_{A'}} \mathbb{V}(X_j) \\ & = \sum_A \sum_{j=J_A+1}^{K_A} \frac{1}{\binom{j}{2}^2} + \sum_{A \cap A' = \emptyset} \sum_{j=J_A \vee J_{A'}+1}^{K_A \wedge K_{A'}} \frac{1}{\binom{j}{2}^2} \\ & \ll \sum_A \frac{1}{J_A^3} + \sum_{A \cap A' = \emptyset} \frac{1}{J_A^3 \vee J_{A'}^3}. \end{aligned}$$

Taking the expectation and using Lemma 4 we obtain that

$$\begin{aligned} \mathbb{E}\left(\left(\sum_{|A|=i} \mathbf{1}_{\mathcal{F}_A} (L_A - \mathbb{E}(L_A \mid \mathcal{G}))\right)^2\right) & \leq \sum_A \sum_{1 \leq j < n} \frac{1}{j^3} \cdot \frac{2j}{(n-1)} \cdot \frac{1}{\binom{n}{i}} \\ & \quad + \sum_{A \cap A' = \emptyset} \sum_{1 \leq j < n} \sum_{j'=1}^j \frac{1}{j^3} \cdot \frac{4jj'}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i, i, n-2i}} \\ & \quad + \sum_{A \cap A' = \emptyset} \sum_{1 \leq j < n} \sum_{j' \geq j+1} \frac{1}{j'^3} \cdot \frac{4jj'}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i, i, n-2i}} \\ & \ll \frac{1}{n}. \end{aligned}$$

This yields (77).

Recall from (76) that $\mathbb{E}(L_A \mid \mathcal{G}) = \frac{2}{J_A} - \frac{2}{K_A}$. In the next step we show that the terms $\frac{2}{K_A}$, $A \in \mathcal{P}_n$, from this representation can be neglected when investigating the evolution of the external length. Using (77), this amounts to saying that for $i \geq 1$ the following claim holds:

$$\sum_{|A|=i} \mathbf{1}_{\mathcal{F}_A} (L_A - \mathbb{E}(L_A)) = \sum_{|A|=i} \mathbf{1}_{\mathcal{F}_A} \left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right)\right) + O_P(n^{-1/2}). \quad (78)$$

In order to see this it suffices to show that

$$\mathbb{E}\left(\left(\sum_{|A|=i} \mathbf{1}_{\mathcal{F}_A} \left(\frac{2}{K_A} - \mathbb{E}\left(\frac{2}{K_A}\right)\right)\right)^2\right) \ll \frac{1}{n}. \quad (79)$$

The claim holds for $i = 1$ since by definition $K_A = n$ for all sets with cardinality one. In the case $i \geq 2$ remember that if $A \cap A' \notin \{A, A', \emptyset\}$ then there cannot be two branches in the coalescent, one supporting the leaves with labels in A and one the leaves with labels in A' and therefore in such case K_A or $K_{A'}$ is (by definition) equal to ∞ and thus $\text{COV}(\frac{2}{K_A}, \frac{2}{K_{A'}}) \leq 0$.

Using Lemma 4 we obtain for $i \geq 2$ that

$$\begin{aligned}
& \mathbb{E}\left(\left(\sum_{|A|=i} \mathbf{1}_{\mathcal{F}_A} \left(\frac{2}{K_A} - \mathbb{E}\left(\frac{2}{K_A}\right)\right)\right)^2\right) \\
& \leq \sum_{|A|=i} \mathbb{V}\left(\frac{2}{K_A}\right) + \sum_{\substack{|A|, |A'|=i \\ A \cap A' = \emptyset}} \mathbb{E}(\mathbf{1}_{\mathcal{F}_A} \mathbf{1}_{\mathcal{F}_{A'}}) \cdot \text{COV}\left(\frac{2}{K_A}, \frac{2}{K_{A'}}\right) \\
& \ll \sum_{|A|=i} \sum_{k=2}^n \frac{1}{k^2} \mathbb{P}(K_A = k) \\
& \quad + \sum_{A \cap A' = \emptyset} \mathbb{E}(\mathbf{1}_{\mathcal{F}_A} \mathbf{1}_{\mathcal{F}_{A'}}) \sum_{k=2}^n \sum_{k'=2}^n \frac{1}{kk'} (\mathbb{P}(K_A = k, K_{A'} = k') - \mathbb{P}(K_A = k) \mathbb{P}(K_{A'} = k')) \\
& \ll \sum_{|A|=i} \sum_{k=2}^n \frac{1}{k^2} \mathbb{P}(K_A = k) + \sum_{A \cap A' = \emptyset} \sum_{k=2}^n \sum_{k'=2}^n \frac{1}{kk'} \mathbb{P}(K_A = k) \mathbb{P}(K_{A'} = k') \cdot \frac{1}{n} \\
& \ll \sum_{|A|=i} \sum_{k=2}^n \frac{1}{k^2} \cdot \frac{k^2}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i}} + \sum_{A \cap A' = \emptyset} \sum_{k=2}^n \sum_{k'=2}^n \frac{1}{kk'} \cdot \frac{k^2 k'^2}{(n-1)^2 (n-2)^2} \cdot \frac{1}{\binom{n}{i}^2} \cdot \frac{1}{n} \\
& \ll \frac{1}{n}
\end{aligned}$$

and thus (79) holds. Putting now together (73) and (78) we obtain the first claim of the proposition.

The calculations above remain valid if we replace $\mathbf{1}_{\mathcal{F}_A}$ by 1. Thus, in the same manner, we also obtain the second claim. \square

6.2 From dependent to independent coefficients

This subsection contains a key building block of the proof of Theorem 1, namely we show that the dependent random coefficients $\{\mathbf{1}_{\mathcal{F}_A}\}_{A \in \mathcal{P}_n}$ which appear in the sum on the right-hand side of (71) can be replaced in the infinite population size limit by random coefficients that are independent of one another and of the coalescent at time 0.

We fix $r \in \mathbb{N}$ and group the branches of the coalescent at time 0 which support at most r leaves (and thus the sets $A \in \mathcal{P}_n, 1 \leq |A| \leq r$) according to whether their initial and final levels are above or below level b_n with

$$b_n := \left\lceil \frac{n}{\log n} \right\rceil.$$

Recall that if there exists no branch in the coalescent at time 0 supporting the leaves with labels in a set A then the initial and final levels K_A and J_A are both equal to ∞ .

Recall also that $\mathbf{1}_{\mathcal{F}_A} = \mathbf{1}_{\{Z_h^A=1\}}$. If the branch supporting the leaves with labels in the set A with $|A| = i$ is formed above level b_n , i.e. $K_A < b_n$, or ends below or at level b_n , i.e. $J_A \geq b_n$, we replace the coefficient $\mathbf{1}_{\mathcal{F}_A}$ by $\widetilde{W}_h^A := \mathbf{1}_{\{\widetilde{B}_h^A=1\}}$, where the processes \widetilde{B}^A are independent birth and death processes started with i individuals at time 0 with birth and death rates equal to $\frac{1}{2}k$ when the process is in state k . Moreover, we choose the birth and death processes to be independent of one another and independent of the coalescent at time 0 and its evolution forwards in time.

Let us now look at the branches that are formed below level b_n and end above this level. These are the b_n branches extant at level b_n . Observe that they support leaves with labels in *disjoint* sets. Out of

these random sets, let $A_1, \dots, A_{l_n}, l_n \leq b_n$, denote those which have cardinalities less than or equal to r : $|A_l| =: i_l \leq r, 1 \leq l \leq l_n$. Let

$$\bar{A} := \bigcup_{l=1}^{l_n} A_l \quad \text{and} \quad \bar{i} := i_1 + \dots + i_{l_n}$$

and note that

$$\bar{i} \leq r \cdot \left\lceil \frac{n}{\log n} \right\rceil. \quad (80)$$

Now we describe our coupling construction. Given the σ -algebra \mathcal{G} , we couple, like in Section 3, independent birth and death processes $\{B^j\}_{j \in \bar{A}}$ to the processes $\{Z^j\}_{j \in \bar{A}}$. In particular, the processes $\{B^{A_l}\}_{1 \leq l \leq l_n}$ are coupled to the processes $\{Z^{A_l}\}_{1 \leq l \leq l_n}$ (the random time change depending on the set \bar{A}). This gives rise to the random variables $W_h^{A_l} := \mathbf{1}_{\{B_h^{A_l}=1\}}$ which we use to replace the coefficients $\mathbf{1}_{\mathcal{F}_{A_l}}$. Since the sets A_1, \dots, A_{l_n} are all disjoint, it follows by Lemma 1 that the processes $B^{A_1}, \dots, B^{A_{l_n}}$, and thus the random variables $W_h^{A_1}, \dots, W_h^{A_{l_n}}$, are independent of one another and independent of the coalescent at time 0.

Through this procedure we obtain the collection of random coefficients

$$V_h^A := \mathbf{1}_{\{J_A < b_n \leq K_A\}} \cdot W_h^A + \mathbf{1}_{\{K_A < b_n\} \cup \{J_A \geq b_n\}} \cdot \widetilde{W}_h^A, \quad A \in \mathcal{P}_{n,r}, \quad (81)$$

where $\mathcal{P}_{n,r} := \{A \subset \{1, \dots, n\}, 1 \leq |A| \leq r\}$. Note that from construction the random variables V_h^A are Bernoulli($p_{i,h}$)-distributed if the set A has cardinality i with $p_{i,h} := \mathbb{P}(B_h^A = 1 \mid B_0^A = i)$. These are the coefficients we use in order to replace the $\mathbf{1}_{\mathcal{F}_A}$'s. In what follows we show that they are independent from one another and independent of the coalescent tree at time 0 (even though J_A and K_A are used in their definition).

Since the dependence on the coalescent tree at time 0 can come only from the tree structure (and not from the exponential times) and the whole information on the tree structure is contained in the σ -algebra \mathcal{G} defined in (75), for the latter property we only need to prove the independence of the collection $\{V_h^A\}_{A \in \mathcal{P}_{n,r}}$ of the σ -algebra \mathcal{G} . To this aim consider bounded, measurable functions $\{f_A\}_{A \in \mathcal{P}_{n,r}}$ depending on the sets A . Then

$$\begin{aligned} & \mathbb{E} \left[\prod_{A \in \mathcal{P}_{n,r}} f_A(V_h^A) \mid \mathcal{G} \right] \\ &= \mathbb{E} \left[\prod_{A \in \mathcal{P}_{n,r}} \left(\mathbf{1}_{\{J_A < b_n \leq K_A\}} \cdot f_A(W_h^A) + \mathbf{1}_{\{K_A < b_n\} \cup \{J_A \geq b_n\}} \cdot f_A(\widetilde{W}_h^A) \right) \mid \mathcal{G} \right] \\ &= \mathbb{E} \left[\sum_{A \in \mathcal{P}(\mathcal{P}_{n,r})} \left(\prod_{A \in \mathcal{A}} (\mathbf{1}_{\{J_A < b_n \leq K_A\}} \cdot f_A(W_h^A)) \cdot \prod_{A \notin \mathcal{A}} (\mathbf{1}_{\{K_A < b_n\} \cup \{J_A \geq b_n\}} \cdot f_A(\widetilde{W}_h^A)) \right) \mid \mathcal{G} \right] \\ &= \sum_{A \in \mathcal{P}(\mathcal{P}_{n,r})} \left(\prod_{A \in \mathcal{A}} \mathbf{1}_{\{J_A < b_n \leq K_A\}} \cdot \prod_{A \notin \mathcal{A}} \mathbf{1}_{\{K_A < b_n\} \cup \{J_A \geq b_n\}} \cdot \mathbb{E} \left[\prod_{A \in \mathcal{A}} f_A(W_h^A) \cdot \prod_{A \notin \mathcal{A}} f_A(\widetilde{W}_h^A) \mid \mathcal{G} \right] \right), \end{aligned}$$

where the last equality follows from the fact that the indicator functions are \mathcal{G} -measurable.

Recall now that if there are two non-disjoint sets $A_1, A_2 \in \mathcal{A}$, then $\mathbf{1}_{\{J_{A_1} < b_n \leq K_{A_1}\}} \cdot \mathbf{1}_{\{J_{A_2} < b_n \leq K_{A_2}\}} = 0$ and therefore the first product in the parenthesis on the right-hand side of the above equality is equal to 0. Using this observation together with the independence of the random variables W_h^A for disjoint sets A , the independence of the random variables \widetilde{W}_h^A and their independence of the σ -algebra \mathcal{G} we obtain that

$$\begin{aligned} \mathbb{E} \left[\prod_{A \in \mathcal{P}_{n,r}} f_A(V_h^A) \mid \mathcal{G} \right] &= \sum_{\substack{A \in \mathcal{P}(\mathcal{P}_{n,r}) \\ \mathcal{A} \text{ contains only disjoint sets}}} \prod_{A \in \mathcal{A}} \mathbf{1}_{\{J_A < b_n \leq K_A\}} \cdot \prod_{A \in \mathcal{A}} \mathbb{E}[f_A(W_h^A)] \\ &\quad \cdot \prod_{A \notin \mathcal{A}} \mathbf{1}_{\{K_A < b_n\} \cup \{J_A \geq b_n\}} \cdot \prod_{A \notin \mathcal{A}} \mathbb{E}[f_A(\widetilde{W}_h^A)]. \end{aligned}$$

Now we can reverse these steps and get that

$$\begin{aligned}\mathbb{E}\left[\prod_{A \in \mathcal{P}_{n,r}} f_A(V_h^A) \mid \mathcal{G}\right] &= \prod_{A \in \mathcal{P}_{n,r}} \mathbb{E}\left[\left(\mathbf{1}_{\{J_A < b_n \leq K_A\}} \cdot f_A(W_h^A) + \mathbf{1}_{\{K_A < b_n\} \cup \{J_A \geq b_n\}} \cdot f_A(\widetilde{W}_h^A)\right)\right] \\ &= \prod_{A \in \mathcal{P}_{n,r}} \mathbb{E}[f_A(V_h^A)]\end{aligned}$$

which proves the independence of the collection $\{V_h^A\}_{A \in \mathcal{P}_{n,r}}$ of the σ -algebra \mathcal{G} and thus of the coalescent tree at time 0. Taking now the expectation in the equality above we also obtain the independence of the random variables V_h^A from one another.

We next show that in the infinite population size limit one can replace the coefficients $\mathbf{1}_{\mathcal{F}_A}$ by the random variables V_h^A , namely that the following holds:

$$\sum_{|A| \leq r} \mathbf{1}_{\mathcal{F}_A} \cdot \left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right)\right) = \sum_{|A| \leq r} V_h^A \cdot \left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right)\right) + O_P\left(\left(\frac{\log \log n}{n}\right)^{1/2}\right). \quad (82)$$

We consider the three groups of branches (for which $K_A < b_n$, $J_A \geq b_n$ and respectively $J_A < b_n \leq K_A$) separately. For each $1 \leq i \leq r$ fixed we have due to independence that

$$\begin{aligned}\mathbb{E}\left(\left(\sum_{|A|=i} (\mathbf{1}_{\mathcal{F}_A} - V_h^A) \cdot \left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right)\right) \cdot \mathbf{1}_{\{K_A < b_n\}}\right)^2\right) \\ = \sum_{|A|=|A'|=i} \mathbb{E}\left((\mathbf{1}_{\mathcal{F}_A} - V_h^A)(\mathbf{1}_{\mathcal{F}_{A'}} - V_h^{A'})\right) \\ \cdot \mathbb{E}\left(\left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right)\right)\left(\frac{2}{J_{A'}} - \mathbb{E}\left(\frac{2}{J_{A'}}\right)\right)\right); K_A, K_{A'} < b_n).\end{aligned}$$

Remember that if $A \cap A' \notin \{A, A', \emptyset\}$ then at least one of the branches with leaves in A , respectively A' , does not exist and by definition $K_A = \infty$ or $K_{A'} = \infty$. Therefore the event $\{K_A, K_{A'} < b_n\}$ is empty and thus in the sum above the summands for which $A \cap A' \notin \{A, A', \emptyset\}$ are equal to 0. We obtain that

$$\begin{aligned}\mathbb{E}\left(\left(\sum_{|A|=i} (\mathbf{1}_{\mathcal{F}_A} - V_h^A) \cdot \left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right)\right) \cdot \mathbf{1}_{\{K_A < b_n\}}\right)^2\right) \\ \leq \sum_A \mathbb{E}\left(\frac{1}{J_A^2} + \mathbb{E}\left(\frac{1}{J_A}\right)^2\right); K_A < b_n) + \sum_{A \cap A' = \emptyset} \left(\mathbb{Cov}\left(\frac{1}{J_A} \cdot \mathbf{1}_{\{K_A < b_n\}}, \frac{1}{J_{A'}} \cdot \mathbf{1}_{\{K_{A'} < b_n\}}\right)\right)^+.\end{aligned} \quad (83)$$

Without loss of generality we can choose $A := \{1, \dots, i\}$ and $A' := \{i+1, \dots, 2i\}$. Let us write for short $K := K_A$, $J := J_A$, $K' := K_{A'}$ and $J' := J_{A'}$.

Note that $K_A < b_n$, respectively $K_{A'} < b_n$ implies that $i \geq 2$, respectively $i' \geq 2$, and thus from Lemma 4

$$\begin{aligned}\mathbb{Cov}\left(\frac{1}{J_A} \cdot \mathbf{1}_{\{K_A < b_n\}}, \frac{1}{J_{A'}} \cdot \mathbf{1}_{\{K_{A'} < b_n\}}\right) \\ = \sum_{k, k' \leq b_n} \sum_{\substack{1 \leq j \leq k-1 \\ 1 \leq j' \leq k'-1}} \frac{1}{jj'} \cdot (\mathbb{P}(K = k, J = j, K' = k', J' = j') - \mathbb{P}(K = k, J = j)\mathbb{P}(K' = k', J' = j')) \\ \leq c \cdot \frac{1}{n} \sum_{k, k' \leq b_n} \sum_{\substack{1 \leq j \leq k-1 \\ 1 \leq j' \leq k'-1}} \frac{1}{jj'} \cdot \mathbb{P}(K = k, J = j)\mathbb{P}(K' = k', J' = j')\end{aligned}$$

$$\begin{aligned}
&\leq c \cdot \frac{1}{n} \sum_{k, k' \leq b_n} \sum_{\substack{1 \leq j \leq k-1 \\ 1 \leq j' \leq k'-1}} \frac{1}{jj'} \cdot \frac{jj'}{(n-1)^2(n-2)^2} \cdot \frac{1}{\binom{n}{i}^2} \\
&\leq c \cdot \frac{1}{n} \cdot \frac{1}{\binom{n}{i}^2}
\end{aligned} \tag{84}$$

for $c < \infty$ depending on i and i' .

Therefore using (54) and again Lemma 4 the inequality (83) becomes

$$\begin{aligned}
&\mathbb{E} \left(\left(\sum_{|A|=i} (\mathbf{1}_{\mathcal{F}_A} - V_h^A) \cdot \left(\frac{2}{J_A} - \mathbb{E} \left(\frac{2}{J_A} \right) \right) \cdot \mathbf{1}_{\{K_A < b_n\}} \right)^2 \right) \\
&\ll \binom{n}{i} \sum_{k \leq b_n} \sum_{j=1}^{k-1} \left(\frac{1}{j^2} + \frac{1}{\binom{n}{i}^2} \right) \cdot \mathbb{P}(K = k, J = j) + \binom{n}{i, i, n-2i} \cdot \frac{1}{n} \cdot \frac{1}{\binom{n}{i}^2} \\
&\ll \binom{n}{i} \sum_{k \leq b_n} \sum_{j=1}^{k-1} \frac{1}{j^2} \cdot \frac{2ij}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i}} + \frac{1}{n} \\
&\ll \sum_{k \leq b_n} \frac{i \log k}{n^2} + \frac{1}{n} \\
&\ll \frac{ib_n \log b_n}{n^2} + \frac{1}{n} \\
&\ll \frac{1}{n}.
\end{aligned} \tag{85}$$

For the branches for which the final level J is greater than or equal to b_n we use a similar argument. Again if $A \cap A' \notin \{A, A', \emptyset\}$ then by definition $J_A = \infty$ or $J_{A'} = \infty$ and thus $\mathbb{C}\mathbb{O}\mathbb{V} \left(\frac{1}{J_A} \cdot \mathbf{1}_{\{J_A \geq b_n\}}, \frac{1}{J_{A'}} \cdot \mathbf{1}_{\{J_{A'} \geq b_n\}} \right) \leq 0$. We obtain for $1 \leq i \leq r$ fixed that

$$\begin{aligned}
&\mathbb{E} \left(\left(\sum_{|A|=i} (\mathbf{1}_{\mathcal{F}_A} - V_h^A) \cdot \left(\frac{2}{J_A} - \mathbb{E} \left(\frac{2}{J_A} \right) \right) \cdot \mathbf{1}_{\{J_A \geq b_n\}} \right)^2 \right) \\
&\ll \sum_A \mathbb{E} \left(\frac{1}{J_A^2} + \mathbb{E} \left(\frac{1}{J_A} \right)^2 \right); J_A \geq b_n \Big) + \sum_{A \cap A' = \emptyset} \left(\mathbb{C}\mathbb{O}\mathbb{V} \left(\frac{1}{J_A} \cdot \mathbf{1}_{\{J_A \geq b_n\}}, \frac{1}{J_{A'}} \cdot \mathbf{1}_{\{J_{A'} \geq b_n\}} \right) \right)^+.
\end{aligned}$$

As above, by Lemma 4

$$\begin{aligned}
\mathbb{C}\mathbb{O}\mathbb{V} \left(\frac{1}{J_A} \cdot \mathbf{1}_{\{J_A \geq b_n\}}, \frac{1}{J_{A'}} \cdot \mathbf{1}_{\{J_{A'} \geq b_n\}} \right) &= \sum_{b_n \leq j, j' < n} \frac{1}{jj'} \cdot (\mathbb{P}(J = j, J' = j') - \mathbb{P}(J = j)\mathbb{P}(J' = j')) \\
&\ll \frac{1}{n} \sum_{b_n \leq j, j' < n} \frac{1}{jj'} \cdot \mathbb{P}(J = j)\mathbb{P}(J' = j') \\
&\ll \frac{1}{n} \sum_{b_n \leq j, j' < n} \frac{1}{jj'} \cdot \frac{jj'}{(n-1)^2} \cdot \frac{1}{\binom{n}{i}^2} \\
&\ll \frac{1}{n} \cdot \frac{1}{\binom{n}{i}^2}
\end{aligned} \tag{86}$$

for $c < \infty$ depending on i and i' and thus

$$\mathbb{E} \left(\left(\sum_{|A|=i} (\mathbf{1}_{\mathcal{F}_A} - V_h^A) \cdot \left(\frac{2}{J_A} - \mathbb{E} \left(\frac{2}{J_A} \right) \right) \cdot \mathbf{1}_{\{J_A \geq b_n\}} \right)^2 \right) \tag{87}$$

$$\begin{aligned}
&\ll \binom{n}{i} \sum_{b_n \leq j < n} \frac{1}{j^2} \cdot \frac{2j}{n-1} \cdot \frac{1}{\binom{n}{i}} + \frac{1}{n} \\
&\ll \frac{1}{n} \cdot \log \frac{n}{b_n} + \frac{1}{n} \\
&\ll \frac{1}{n} \cdot \log \log n.
\end{aligned}$$

We are now left to consider the branches that are formed below and end above level b_n in the coalescent. Again, using the same argument as above and the fact that for $A \cap A' \notin \{A, A', \emptyset\}$ the event $\{J_A < b_n \leq K_A, J_{A'} < b_n \leq K_{A'}\}$ is empty (since there can not be two branches in the tree that coexist at level b_n and have common leaves), we have that

$$\begin{aligned}
&\mathbb{E}\left(\left(\sum_{|A|=i} (\mathbf{1}_{\mathcal{F}_A} - V_h^A) \cdot \left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right)\right) \cdot \mathbf{1}_{\{J_A < b_n \leq K_A\}}\right)^2\right) \\
&\ll \sum_A \mathbb{E}\left((\mathbf{1}_{\mathcal{F}_A} - W_h^A)^2\right) \cdot \mathbb{E}\left(\frac{1}{J_A^2} + \mathbb{E}\left(\frac{1}{J_A}\right)^2; J_A < b_n \leq K_A\right) \\
&\quad + \sum_{A \cap A' = \emptyset} \text{COV}\left(\frac{1}{J_A} \cdot \mathbf{1}_{\{J_A < b_n \leq K_A\}}, \frac{1}{J_{A'}} \cdot \mathbf{1}_{\{J_{A'} < b_n \leq K_{A'}\}}\right) \\
&= \sum_A \mathbb{P}\left(\mathbf{1}_{\{Z_h^A=1\}} \neq \mathbf{1}_{\{B_h^A=1\}}\right) \cdot \mathbb{E}\left(\frac{1}{J_A^2} + \mathbb{E}\left(\frac{1}{J_A}\right)^2; J_A < b_n \leq K_A\right) \\
&\quad + \sum_{A \cap A' = \emptyset} \text{COV}\left(\frac{1}{J_A} \cdot \mathbf{1}_{\{J_A < b_n \leq K_A\}}, \frac{1}{J_{A'}} \cdot \mathbf{1}_{\{J_{A'} < b_n \leq K_{A'}\}}\right). \tag{88}
\end{aligned}$$

If the cardinality i of the sets A and A' is at least 2, then the second sum on the right hand side can be bounded from above by $c \cdot \frac{1}{n}$ for a finite constant c by using the same argument as the one used to obtain (84). If $i = 1$ recall that $K_A = n$ and thus $\{J_A < b_n \leq K_A\} = \{J_A < b_n\}$ (and the same for the set A'). By the argument used to obtain (86) we can again bound the sum from above by $c \cdot \frac{1}{n}$. As to the first sum on the right hand side of (88) recall that we coupled all processes Z^A with A such that $J_A < b_n \leq K_A$ simultaneously and that the random time change depends just on the union \bar{A} of these sets. From (17) and (80) we obtain that

$$\mathbb{P}\left(\mathbf{1}_{\{Z_h^A=1\}} \neq \mathbf{1}_{\{B_h^A=1\}}\right) \ll \frac{\bar{i}}{n} \ll \frac{1}{\log n},$$

which plugged in (88) leads to

$$\begin{aligned}
&\mathbb{E}\left(\left(\sum_{|A|=i} (\mathbf{1}_{\mathcal{F}_A} - V_h^A) \cdot \left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right)\right) \cdot \mathbf{1}_{\{J_A < b_n \leq K_A\}}\right)^2\right) \tag{89} \\
&\ll \frac{1}{\log n} \sum_A \mathbb{E}\left(\frac{1}{J_A^2} + \mathbb{E}\left(\frac{1}{J_A}\right)^2\right) + \frac{1}{n} \\
&\ll \frac{1}{\log n} \binom{n}{i} \sum_{j < n} \left(\frac{1}{j^2} + \frac{1}{\binom{n}{i}^2}\right) \cdot \mathbb{P}(J = j) + \frac{1}{n} \\
&\ll \frac{\binom{n}{i}}{\log n} \sum_{j < n} \frac{1}{j^2} \cdot \frac{2j}{(n-1)} \cdot \frac{1}{\binom{n}{i}} + \frac{1}{n} \\
&\ll \frac{1}{\log n} \cdot \frac{\log n}{n} + \frac{1}{n} \\
&\ll \frac{1}{n}.
\end{aligned}$$

Putting now together (85), (87) and (89) we obtain (82).

6.3 Completion of the proof of the Theorem

Let $t > 0$ be a fixed time measured on the evolutionary time-scale (the scale on which the lengths in the coalescent are measured) and recall that $h = \frac{t}{n}$. Remember from (1) that, considering the coalescent tree of the population alive at time h , $I_{0,h}^n$ denotes the amount of external length in the coalescent at time h that is gathered in the region near the leaves of the tree, more precisely in the time interval $(0, h)$. We start by showing that

$$I_{0,h}^n - \mathbb{E}(I_{0,h}^n) = O_P(n^{-1/2}). \quad (90)$$

Let \tilde{X}_j , $2 \leq j \leq n$, denote the inter-coalescence times in the coalescent of the population alive at time h and consider the random time

$$T := \sum_{j \geq \frac{2}{t+2}n}^n \tilde{X}_j.$$

Since $\tilde{X}_j \sim \text{Exp}\left(\binom{j}{2}\right)$, it holds that $\mathbb{E}(T) = \frac{t}{n} + O\left(\frac{1}{n^2}\right)$ and $\mathbb{V}(T) = O\left(\frac{1}{n^3}\right)$. Observe that the quantity $I_{0,\frac{t}{n}}^n - I_{0,T}^n$ is the external length gathered in an interval of length $|\frac{t}{n} - T|$. In this time interval some of the n external branches may become internal. It follows that

$$|I_{0,\frac{t}{n}}^n - I_{0,T}^n| \leq n \cdot \left| \frac{t}{n} - T \right|$$

and thus

$$I_{0,\frac{t}{n}}^n - \mathbb{E}(I_{0,\frac{t}{n}}^n) = I_{0,T}^n - \mathbb{E}(I_{0,T}^n) + O_P(n^{-1/2}).$$

In order to estimate $I_{0,T}^n$ we use Proposition 3 in [18]. The arguments used in its proof apply completely unchanged also for α and β depending on n . Choosing α such that $n^\alpha = \frac{2}{2+t}n$ and $\beta = 1$ the proof gives

$$\mathbb{V}(I_{0,T}^n) = 8 \left(1 - \frac{\log\left(\frac{2}{t+2}n\right)}{\log n} \right) \cdot \frac{\log n}{n} + O(n^{-1}) = O(n^{-1}),$$

and therefore (90) holds.

We are now ready for the final step. For ease of notation let us now consider just two times $0 < h_1 < h_2$. We show that for any $\alpha_1, \alpha_2 \in \mathbb{R}$ the linear combination

$$\sqrt{\frac{n}{4 \log n}} \left(\alpha_1 (\mathcal{L}_{h_1}^n - \mathbb{E}(\mathcal{L}_{h_1}^n)) + \alpha_2 (\mathcal{L}_{h_2}^n - \mathbb{E}(\mathcal{L}_{h_2}^n)) \right)$$

of the (centred and rescaled) lengths at the two times converges in distribution to a normal distributed random variable. The argument can be immediately extended for linear combinations of the lengths at times $0 < h_1 < \dots < h_k$, $k \in \mathbb{N}$.

Using the notation given at the beginning of Section 5 we have that for each $r \in \mathbb{N}$

$$\mathcal{L}_{h_j}^n = \sum_{|A| \leq r} \mathbf{1}_{\mathcal{F}_A(h_j)} \cdot L_A + \tilde{R}_{h_j}^{n,r} + \bar{R}_{h_j}^{n,r} + I_{0,h_j}^n, \quad j = 1, 2.$$

Here we write $\mathcal{F}_A(h_j)$ for the event that the number of descendants at time h_j , $j \in \{1, 2\}$, of the individuals with labels in the set A at time 0 is equal to 1. The second and third terms have been treated in Proposition 2 and we now focus on the first term. From (71)

$$\sum_{|A| \leq r} \mathbf{1}_{\mathcal{F}_A(h_j)} \cdot L_A - \mathbb{E} \left(\sum_{|A| \leq r} \mathbf{1}_{\mathcal{F}_A(h_j)} \cdot L_A \right) = \sum_{|A| \leq r} \mathbf{1}_{\mathcal{F}_A(h_j)} \left(\frac{2}{J_A} - \mathbb{E} \left(\frac{2}{J_A} \right) \right) + O_P(n^{-1/2}).$$

Equation (82) allows to replace the dependent random coefficients $\mathbf{1}_{\mathcal{F}_A(h_1)}$ by the independent coefficients $V_{h_1}^A$ defined in (81). In order to obtain the corresponding independent random coefficients for the time h_2

we let the birth and death processes $\widetilde{W}_{h_1}^A$ and $W_{h_1}^A$ used to define $V_{h_1}^A$ run further until time h_2 and define the coefficients $V_{h_2}^A$ as in (81) by using the states of the processes $\widetilde{W}_{h_2}^A$ and $W_{h_2}^A$ at time h_2 . Note that the argument from the previous subsection used to prove (82) continues to hold also for the time h_2 and the coefficients defined as above. We thus obtain using (82), (90) and Proposition 2 that for $r \in \mathbb{N}$

$$\begin{aligned} & \sqrt{\frac{n}{4 \log n}} \left(\alpha_1 (\mathcal{L}_{h_1}^n - \mathbb{E}(\mathcal{L}_{h_1}^n)) + \alpha_2 (\mathcal{L}_{h_2}^n - \mathbb{E}(\mathcal{L}_{h_2}^n)) \right) \\ &= \sqrt{\frac{n}{4 \log n}} \sum_{|A| \leq r} \left(\alpha_1 V_{h_1}^A + \alpha_2 V_{h_2}^A \right) \cdot \left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right) \right) \\ &+ \sqrt{\frac{n}{4 \log n}} \left(\alpha_1 (\widetilde{R}_{h_1}^{n,r} - \mathbb{E}(\widetilde{R}_{h_1}^{n,r})) + \alpha_2 (\widetilde{R}_{h_2}^{n,r} - \mathbb{E}(\widetilde{R}_{h_2}^{n,r})) \right) \\ &+ O_P\left(\left(\frac{\log \log n}{\log n}\right)^{1/2}\right). \end{aligned}$$

We now argue that the collection of random variables $\left\{ \sqrt{\frac{n}{4 \log n}} \left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right) \right), 1 \leq i \leq r, |A| = i \right\}$, fulfils the assumptions of Proposition 1. First we make use of the main result of [5] which says that the (centred and rescaled) internal lengths of different orders are asymptotically normal and independent. Together with (72) it follows that

$$\left(\sqrt{\frac{n}{4 \log n}} \sum_{|A|=i} \left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right) \right) \right)_{i=1, \dots, r} \xrightarrow{d} N(0, I_r)$$

where I_r is the $r \times r$ identity matrix. The fact that the other two assumptions of Proposition 1 hold has been proved in Lemma 5. Therefore, we can apply Proposition 1 for the collection of random variables $\left\{ \sqrt{\frac{n}{4 \log n}} \left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right) \right), 1 \leq i \leq r, |A| = i \right\}$, and the independent random coefficients $\{\alpha_1 V_{h_1}^A + \alpha_2 V_{h_2}^A\}, 1 \leq i \leq r, |A| = i$.

Letting now $n \rightarrow \infty$ and using again Proposition 2 we obtain that the linear combination of the external lengths at the times h_1 and h_2 is asymptotically normal distributed. Note that we may interchange the limits in r and n since Proposition 2 gives a uniform estimate in r . Therefore the stationary process \mathcal{L}^n converges in finite distributions to a stationary Gaussian process. It remains to compute its covariance function.

Observe that for $h_1 = 0$ and $h_2 = h$

$$\alpha_1 V_{h_1}^A + \alpha_2 V_{h_2}^A = \alpha_1 \delta_{1, |A|} + \alpha_2 V_h^A$$

with second moment from (14)

$$m_A = \begin{cases} \alpha_1^2 + 2\alpha_1\alpha_2 \left(\frac{2}{h+2}\right)^2 + \alpha_2^2 \left(\frac{2}{h+2}\right)^2, & \text{if } |A| = 1 \\ \alpha_2^2 i \cdot \left(\frac{h}{h+2}\right)^{i-1} \left(\frac{2}{h+2}\right)^2 & \text{if } |A| = i > 1. \end{cases}$$

Therefore from Proposition 1 as $n \rightarrow \infty$

$$\begin{aligned} & \sqrt{\frac{n}{4 \log n}} \sum_{|A| \leq r} \left(\alpha_1 V_0^A + \alpha_2 V_h^A \right) \cdot \left(\frac{2}{J_A} - \mathbb{E}\left(\frac{2}{J_A}\right) \right) \\ & \xrightarrow{d} N\left(0, \alpha_1^2 + 2\alpha_1\alpha_2 \left(\frac{2}{h+2}\right)^2 + \alpha_2^2 \left(\frac{2}{h+2}\right)^2 \sum_{i=1}^r i \cdot \left(\frac{h}{h+2}\right)^{i-1}\right). \end{aligned}$$

Taking the limit $r \rightarrow \infty$ we obtain that

$$\sqrt{\frac{n}{4 \log n}} \left(\alpha_1 (\mathcal{L}_0^n - \mathbb{E}(\mathcal{L}_0^n)) + \alpha_2 (\mathcal{L}_h^n - \mathbb{E}(\mathcal{L}_h^n)) \right) \xrightarrow{d} N\left(0, \alpha_1^2 + 2\alpha_1\alpha_2 \left(\frac{2}{h+2}\right)^2 + \alpha_2^2\right) \quad \text{as } n \rightarrow \infty.$$

This gives the covariance function.

The almost sure continuity of the paths follows from Lemma 6.4.6 in [22] using stationarity. To apply the lemma one has to check that there exists a $\delta > 0$ such that

$$\int_0^\delta \frac{\sigma^+(u)}{u \left(\log \frac{1}{u} \right)^{1/2}} du < \infty,$$

where

$$\sigma^+(u) := \sup_{\substack{|t-s| \leq u \\ t, s \in [-1/2, 1/2]}} (\mathbb{E}(\mathcal{L}_t^1 - \mathcal{L}_s^1)^2)^{1/2}.$$

This criterion is easily verified in our case, which finishes the proof.

References

- [1] Athreya, K.B., Ney, P.E.: Branching Processes. Springer (1972)
- [2] Berestycki, J., Berestycki, N., Limic, V.: Asymptotic sampling formulae for Lambda-coalescents. To appear in Ann. Inst. H. Poincaré, arXiv: 1201.6512 (2012)
- [3] Berestycki, J., Berestycki, N., Schweinsberg, J.: Beta-coalescents and continuous stable random trees. Ann. Probab. 35, 1835-1887 (2007)
- [4] Berestycki, J., Berestycki, N., Schweinsberg, J.: Small time properties of Beta-coalescents. Ann. Inst. H. Poincaré 44, 214-238 (2008)
- [5] Dahmer, I., Kersting G.: The internal branch lengths of the Kingman coalescent. Ann. Appl. Probab. 25, No. 3, 1325-1348 (2015)
- [6] Dahmer, I., Kersting G., Wakolbinger, A.: The total external branch length of Beta-coalescents. Comb. Probab. Comput., Special Issue on Analysis of Algorithms 23, Special Issue 06, 1010-1027 (2014)
- [7] Dahmer, I., Knobloch, R., Wakolbinger, A.: The Kingman tree length process has infinite quadratic variation. Electron. Comm. Probab. 19, 1-12 (2014)
- [8] Depperschmidt, A., Greven, A., Pfaffelhuber, P.: Tree-valued Fleming-Viot dynamics with mutation and selection. Ann. Appl. Probab. 22(6), 2560-2615 (2012)
- [9] Dhersin, J.-S., Yuan, L.: Asymptotic behavior of the total length of external branches for Beta-coalescents. Adv. Appl. Probab. 47, No. 3, 693-714 (2015)
- [10] Drmota, M., Iksanov, A., Möhle, M., Rösler, U.: Asymptotic results about the total branch length of the Bolthausen-Sznitman coalescent. Stoch. Proc. Appl. 117, 1404-1421 (2007)
- [11] Durrett, R.: Probability Models for DNA Sequence Evolution, 2nd ed. Springer, New York (2008)
- [12] Ethier, S.N., Kurtz, T.G.: Markov processes. Characterization and convergence. John Wiley and sons (1986)
- [13] Fu, Y.X.: Statistical properties of segregating sites. Theor. Pop. Biol. 48, 172-197 (1995)
- [14] Greven, A., Pfaffelhuber, P., Winter, A.: Convergence in distribution of random metric measure spaces: (Lambda-coalescent measure trees). Prob. Theo. Rel. Fields 145(1), 285-322 (2009)
- [15] Greven, A., Pfaffelhuber, P., Winter, A.: Tree-valued resampling dynamics. Martingale problems and applications. Prob. Theo. Rel. Fields 155, 789-838 (2013)

- [16] Gufler, S.: Lookdown representation for tree-valued Fleming-Viot processes. arXiv:1404.3682 (2014)
- [17] Iksanov, A., Möhle, M.: A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Comm. Probab.* 12, 28-35 (2007)
- [18] Janson, S., Kersting, G.: On the total external length of the Kingman coalescent. *Electron. J. Probab.* 16, 2203-2218 (2011)
- [19] Kersting, G.: The asymptotic distribution of the length of Beta-coalescent trees. *Ann. Appl. Probab.* 22, 2086-2107 (2012)
- [20] Kersting, G., Schweinsberg, J., Wakolbinger, A.: The evolving beta coalescent. *Electron. J. Probab.* 19, 1-27 (2014)
- [21] Möhle, M.: Asymptotic results for coalescent processes without proper frequencies and applications to the two-parameter Poisson-Dirichlet coalescent. *Stoch. Process. Appl.* 120, 2159-2173 (2010)
- [22] Marcus, M.B., Rosen, J.: *Markov Processes, Gaussian Processes and Local Times*. Cambridge University Press (2006)
- [23] Pfaffelhuber, P., Wakolbinger, A.: The process of most recent common ancestors in an evolving coalescent. *Stoch. Processes Appl.* 116, 1836-1859 (2006)
- [24] Pfaffelhuber, P., Wakolbinger, A., Weisshaupt, H.: The tree length of an evolving coalescent. *Prob. Theo. Rel. Fields* 151, 529-557 (2011)
- [25] Schweinsberg, J.: Dynamics of the evolving Bolthausen-Sznitman coalescent. *Electron. J. Probab.* 91, 1-50 (2012)

